Investigating the drivers of regional variation in tonsillectomy rates and patient and surgeon preference elicitation in treatment choice of adults with recurring tonsillitis

Dr Nishchay Mehta

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

UCL Farr Institute of Health Informatics Research

UCL Ear Institute, evidENT

Supervisors:

Professor Anne Schilder

Professor Andrew Hayward

Dr Sarah Smith

Declaration

I, Dr Nishchay Mehta, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Candidate’s signature

Thesis Abstract

Introduction

Rates of surgery vary considerably across the UK. Many assume that this on the one hand exposes residents of certain UK regions to unnecessary surgical risks, and on the other hand prevents those of neighbouring regions from receiving important surgical care. Nowhere is this more apparent than for tonsillectomy – an operation that involves removing the tonsils in patients suffering with severe recurring sore throats. With 40,000 tonsillectomies per year, it is one of the most common operations in the NHS, but remarkably, tonsillectomies are done seven times more frequently in some UK regions than in others. Despite national efforts to reduce these differences (e.g. re-education programmes for ENT surgeons, creation of national guidance on how to manage recurring sore throat, and financial penalties locally imposed to restrict high numbers of surgeries) this disparity has only got worse over the past 17 years. I undertook my research to examine the causes of these differences in more detail, to guide future policies.

Aims

The aim of my thesis was to develop a better understanding of the drivers of regional tonsillectomy rate variation by quantifying regional variation of tonsillectomy rates in relation to regional demands, and by exploring the role of professional uncertainty and treatment preference on the treatment chosen.

Objectives

The objectives were to establish the:

A. Rate and regional variation of self-reported sore throat and help seeking behaviour in the community;
B. Rate and regional variation of recurring sore throat in primary care;
C. Rate and regional variation of tonsillectomy in secondary care, after adjusting for local rates of recurring sore throat;
D. Constructs of clinical decision making and thereby ascertain which concepts were most likely to be related to surgical rate variation;
E. Role of surgeon and patient decisional uncertainty on the treatment chosen for recurring tonsillitis;

F. Role of surgeon and patient treatment preference on the treatment chosen for recurring tonsillitis;

Methods

By using the largest UK population based study of upper respiratory symptoms and primary-secondary care linked medical record databases, I was able to investigate regional surgical rate variations across the entire patient-disease pathway: from sore throat in the community (used as a surrogate marker for tonsillitis), through recurrent sore throat consultations in primary care and finally tonsillectomy in hospitals. Following systematic review and thematic analysis of patient focus groups, I designed an instrument to elicit patient and surgeon preference. By undertaking the largest observational study of decision-making in adults with recurring tonsillitis, I was able to investigate the role of both patient and surgeon treatment preference and decisional uncertainty on treatment choice.

Results

My results suggest:

A. There is considerable variation in the incidence of self-reported severe sore throat between regions. However, once patient risk factors are accounted for there is no statistical evidence for disparity between regions. In those who self-report a severe sore throat there is also a degree of regional variation in the rate of relevant consultations for sore throat symptoms, however, once disease characteristics were accounted for, this regional disparity disappears.
B. There is regional disparity in recurring sore throat consultations in primary care, however, once patient characteristics are accounted for, this regional variation reduces considerably.
C. Similarly, there is regional disparity in tonsillectomy rates; this variation reduces considerably once patient characteristics are accounted for.
D. In the literature concepts related to shared-decision-making are strongly inter-related and often poorly defined. Decisional uncertainty and treatment preference are amongst the best described, most measurable, and most appropriate constructs to investigate in a study of surgical rate variation.

E. Decisional uncertainty, either patient's or surgeon's, was found to have no role to play in the treatment chosen during a consultation for recurring sore throat.
F. Patients' treatment preferences did not influence their treatment chosen, but surgeons' treatment preferences did.

Discussion

There are three key findings of my thesis. Firstly, regional rate of consultation for sore throat – which was used as a proxy for the underlying tonsillitis rate in the population throughout – was mirrored in the regional rate of tonsillectomy. This implies the regional tonsillectomy rate variations reflect regional variations in the 'need' of the population. Secondly, regional tonsillectomy rate variations are greater for children than adults. Finally, treatment decisions for adults with recurring tonsillitis are more influenced by surgeon's treatment preferences than patient preferences or severity.

There is a strong culture within the NHS of addressing variations of all kinds as a means of increasing healthcare quality and decreasing cost. There are currently metrics of variation across almost every aspect of care, however few of these account for patient characteristics to the extent that this thesis has, meaning that the initiatives may be a waste of effort at best and harmful at worst.

The work presented in my thesis uses a unique set of mixed methods to demonstrate the complexity of regional tonsillectomy rate variation, which too frequently has been investigated using poorly controlled cross-sectional studies and reduced to soundbites like the "Surgical Signature". Whilst my study shows "surgical signature" is important, it fails to describe the true complexity of the variation observed. My study sheds more light on the complexity of this variation and provides a plausible reason as to why the policies to reduce tonsillectomy rate variations may have failed. This mixed methods approach could be used more broadly to inform discussions under regional surgical rate variations. Most importantly, the findings in this thesis also demonstrate where future policy could be targeted to reduce unwarranted regional tonsillectomy rate variation.

Acknowledgements

Academic support

I would like to thank my PhD supervisors Prof. Andrew Hayward, Dr Sarah Smith and Prof. Anne Schilder for their advice and support in preparing my fellowship application and during my PhD. Whilst each supervisor had input into all chapters of this thesis, day-to-day guidance on epidemiological studies (chapters 2,3 and 4) was undertaken by Prof Hayward, whilst Dr Smith provided the same level of support for work on my chapters related to decision making (chapters 5, 6 and 7). I would also like to thank Professor Alessio Ishizaka for supervising my preference elicitation work using Analytical Hierarchical Process. Professor Ishizaka guided the creation of PARTT and analysis of DART preference results. Professor Schilder provided overall oversight to the project and helped me guide my overall discussions.

I would also like to thank my colleagues at UCL: Dr Hannah Evans, who advised on the statistical components of my analysis, Dr Spiros Denaxas who provided me with a strong foundation in CALIBER, Dr Logan Manikan who helped as a second reviewer for phenotyping conditions and systematic reviews, Aneeka Deegun who helped run focus groups, Giovanna Ceroni, Cherry Heaven, Kim Airey and Dr Helen Blackshaw who provided administrative support throughout the PhD.

I especially want to thank my wife Karen Mehta, and my two children Maya and Seth, who have been amazingly supportive and encouraging throughout the whole PhD.

Financial support

I am in receipt of a Wellcome Trust Research Training Fellowship (WT101675AIA) hosted at the UCL Ear Institute. FluWatch is a collaboration between epidemiologists at the Centre for Infectious Disease Epidemiology at University College London (UCL), virologists and mathematical modelers from the Health Protection Agency (HPA, now Public Health England), immunologists at the Medical Research Council (MRC) Human Immunology Unit at Oxford University and the MRC General Practice Research Framework (GPRF). Funded by the MRC it began recruitment in 2006, however, when the H1N1 pandemic arose in 2009 further funding was secured jointly from the MRC and Wellcome Trust, allowing continued follow-up and an expansion in cohort size.

CALIBER is a collaboration between UCL, the London School of Hygiene and Tropical Medicine (LSHTM) and the Clinical Practice Research Datalink (CPRD). CALIBER is funded by a Wellcome Trust project grant (086091/Z/08/Z) and a NIHR programme grant (RP-PG-0407-10314).

This study was supported by the UCL Farr Institute of Health Informatics Research by providing both infrastructure, access and technical support in utilising CALIBER.

Patient perspective

The inception of this project was informed by concerns from patients with recurring tonsillitis who were convened for Patient Public Involvement process of a tonsillectomy trial application. The patients reported great variation in treatments for recurring tonsillitis and suggested studying why the process of receiving a tonsillectomy varied so greatly.

Personal Perspective

Below I present some reflections of my personal experience during the PhD. Since many key decisions are made before the PhD even starts, they are not always made explicit in the chapters that follow. They relate to the weeks of deliberate thinking before a methodology becomes a study, when other potential methodologies are buried and forgotten. I found these decisions to be amongst the most difficult to take and convey in the thesis. At the inception of my PhD I had weekly meetings with Dr Sarah Smith and Prof Andrew Hayward, where we discussed the aims and potential methodological approaches that I had explored through MSc modules in social epidemiology or through private reading. These were supplemented with 3 weekly “all supervisors” meetings where I presented the potential strengths and weaknesses of different methodologies. I also had the advantage of presenting my plans to supervisors on the epidemiology modules I undertook and the conferences I attended. These conversations helped me centre my investigations so they could be informative, but also provide future options to effect a change.

I began this journey because my personal experience of working in different regions of the country as a surgical trainee did not agree with the strongly held academic view that there is a strong ‘surgical signature’ effect that results in high tonsillectomy rates in some regions and not in others. So this was an extremely personal journey that began from my

personal experience, but required me to acknowledge and address the biases resulting from personal experiences to really investigate the issues.

When I reviewed the evidence base, I noted that studies had rarely accounted for differences in underlying disease burden between regions and whilst professional uncertainty and surgeon preference were thought to be the main drivers of variation, there was little proof that these factors influenced medical decisions.

An ideal thesis would have allowed me to investigate these variations at all levels of the patient pathway, from tonsillitis in the community, to GP consultations for tonsillitis and recurring tonsillitis, outpatient referrals for recurring tonsillitis and tonsillectomies at hospital and local health authority policies for tonsillectomy. Additionally, an investigation into decision making for treatments of recurring tonsillitis, would have allowed me to look at the decision-making from multiple visits in primary care through to the surgeon-patient interaction when the final treatment decision was made.

However, tonsillitis has no objective test and was poorly diagnosed and coded in primary care records. , Additionally, patients may experience tonsillitis but not recognise their illness as such or even consult for their symptoms.. There was considerable heterogeneity in the codes used to diagnose tonsillitis in primary care and poor coding for specialist referrals for recurring tonsillitis. Additionally, due to data safety regulations rules around re-identification of pseudo-anonymised data, analysing patient data mapped to hospitals and local health care trusts was not permitted. Finally, a study of decision making in primary care where only a small proportion of consultations relate to tonsillitis would have required ineffective resource allocation over a considerable length of time.

Therefore, pragmatically, I chose to investigate severe sore throat in the community and in primary care, instead of tonsillitis. Using a sensitive definition of my cases (and treating sore throat as a surrogate marker of tonsillitis) ensured that I did not misinterpret variations in the manner people define and diagnose tonsillitis with variations in the incidence of tonsillitis. Whilst I could not investigate variations at the level of the hospital due to data safety restrictions, I was able to investigate them at the level of 10 larger health care regions. This allowed me a first step to developing a better understanding of tonsillectomy rate variation, without which, future studies would not be able to request information on hospital identities.

Since the role of decision-making in recurring tonsillitis had not been investigated directly at all in relation to regional tonsillectomy rate variation, I felt a well-targeted large observational study would be more valuable than a small qualitative study, in the first instance, especially if the goal was to inform on regional tonsillectomy rate variation. It would highlight areas that could be targeted more directly in future studies. Most of the evidence concludes that surgical signature is responsible for regional variation in tonsillectomy rate. I decided to study decision making during the consultation between ENT surgeon and patient. However, it was still unclear what I should study during the shared decision making process that occurs during that consultation and so I had to unpack the conceptual frameworks of shared decision making with the aim of finding a concept that was most likely related to regional surgical rate variation. I chose decisional uncertainty and surgeon and patient preference as they were most likely related to regional tonsillectomy rate variation, compared to other conceptual frameworks, based on the available, but weak, evidence. This decision was based on the assumption that uncertain surgeons are more likely to base their decisions on their personal preferences; therefore, if the assumption was not valid it may have led to uninformative findings. However, if the conceptual framework I identified was related to regional tonsillectomy rate variation it would be more amenable to a targeted intervention to improve the situation.

Finally, there was no specific instrument available to elicit preferences in the treatment of recurring tonsillitis. Rather than simplifying preference for surgery into a visual analogue scale that lacked validity – which had been done frequently in the realm of orthopaedic surgery - I chose to develop a new instrument based on the available evidence. I generated this instrument based on a systematic review and qualitative analysis of patient focus groups. This took me considerably longer, but I believe it allowed me to create an instrument that could be used in the future, based on a more valid representation of the underlying construct being measured.

Overall, I gained an in-depth knowledge of data management, advanced epidemiology, analysis of person level data and multivariable modelling, decision analysis, systematic reviews and an introduction into qualitative methods. I was the Chief Investigator in the largest multi-centre portfolio study of surgical decision making to date and gained invaluable experiences in this sphere.

Additional short courses

- Multilevel modelling for health research, Department of Epidemiology, Graduate School, UCL 02-04/04/2014
- Focus Groups in Qualitative Research, Department of Primary Care and Population health UCL – Theory and Practice 05/03/13
- Qualitative data analysis, Department of Primary Care and Population health UCL 12/03/13
- Introduction to Data Linkage, Administrative Data Research Centre for England, Faculty of Social and Human Sciences, University of Southampton 12/09/2014
- An Introduction to Primary Care Databases & Missing Data abd New Methods of Multiple Imputation of Longitudinal Electronic Health Records, Department of Primary Care and Population Health, UCL 11 – 15/11/ 2013
- Advanced Analysis of Linked Health Data course, Helath Informatics Group, College of Medicine, Institute of Life Sciences, Swansea University 23-27/03/14
- Multi Criteria Decision Analysis Maser Class, Business Services and Research Office, Univesity of Portsmouth Business School 05/06/2015

Conferences attended

- Student fellow, International Society Quality of Life Annual Conference, Miami, October 2013. Attended worskhops on basic principles of psychometrics, health related quality of life meaure and how to create a questionnaire. Discussed upcoming PhD with Professor John Browne (Professor of Psychometrics LSHTM), who described the difficulties of measuring preferences reliably.
- Student fellow,International Society Quality of Life Annual Conference, Berlin, October 2014. Attended workshops on preference elicitation using discrete choice experiments and discussed preference measurement with Prof Axel Muelbacher (professor of preference and decision analysis) and Professor John Brazier (designer of SF36). Both encouraging of work and of robust method being used to develop questionnaire.

- Student fellow, International Society Quality of Life Annual Conference, Vancouver, October 2015. Learnt about new techniques in advanced psychometrics.
- Student, International Symposium of Analytical Hierarchical Conference, London, 04-07/08/2016. Discussed AHP analysis with William Adams (AHP expert and software designer) who suggested collaboration to create on line portal to deliver questionnaire.

Presentations

- **Regional tonsillectomy rate variation.** Otorhinololaryngological Research Society, Royal College of Surgeons 13/03/15: In this forum I have presented emerging methods and findings used in the FluWatch Study
- **Are ENT surgeons responsible for regional tonsillectomy rate variations?** Dutch ENT Society 21/04/2015. In this forum, I have presented findings used in the FluWatch and CALIBER studies
- **Decision Making in Adults with Recurring Tonsillitis** Royal National Throat Nose and Ear Hospital 19/05/2015, In this forum I have presented emerging methods and findings from the Decision Making study
- **Investigating drivers of regional tonsillectomy rate variation,** Farr Institute, Upgrade presentation 14/10/2015, In this forum I have presented emerging methods and findings used in the FluWatch and CALIBER and decision-making studies
- **Using electronic health care records to investigate healtcare variaitons in tonsillectomy rates** British Society of Academic Otorhinolaryngologists 10/03/2016, In this forum I have presented emerging methods used in the FluWatch and CALIBER studies
- **Understanding surgical rate variations in across patient pathways**, Grand Round, Royal National Throat Nose and Ear Hospital 27/01/2016, In this forum I have presented emerging methods and findings used in the FluWatch and CALIBER and decision-making studies
- **Making sense of medical literature** Pan Thames Research Lectures for ENT trainees 08/12/2016, In this forum I have presented emerging methods and findings of decision-making studies

- **Drivers of regional variations in tonsillectomy rates** ENTUK Annual Conference, York, 31/03/2017, In this forum I have presented emerging methods and findings used in the FluWatch and CALIBER and decision-making studies
- **Regional variation in local health care policy** Hartopp-Dixon Annual Presentations 08/07/2017. In this forum I have presented findings from my study of local policy regarding tonsillectomy

Prizes

- Hartropp Dixon Prize 08/07/2017, Yearly competition for Junior Doctors in ENT. **Regional variation in local health care policy** In this forum I have presented findings from my study of local policy regarding tonsillectomy (Cash Prize £250)

- Phillip Stell Prize at Otorhinolaryngological Research Society 13/03/2015, Bi-annual competition for ENT surgeons. In this forum, I have presented emerging methods and findings used in the FluWatch Study. Awarded costs to present to Dutch Society (up to £500)

Publications undertaken during PhD fellowship

- **Chapter 30: Ear, Nose Throat and Eye Diseases.** Vaz F, Mehta N. Clinical Medicine, Ninth Edition, 2017, Elsevier Publishing. Not directly related to PhD but important to professional development as working on teaching and writing skills.

- **Antibiotic prescribing in patients with self-reported sore throat.** Mehta N, Schilder A, Fragaszy E, E R Evans H, Dukes O, Manikam L, Little P, Smith SC, Hayward A. J Antimicrob Chemother. 2017;72(3):914-922. Directly related to FluWatch study, methods and results.

- **Shared Decision Making and Choice for Elective Surgical Care: A Systematic Review.** Boss EF, Mehta N, Nagarajan N, Links A, Benke JR, Berger Z, Espinel A, Meier J, Lipstein EA. Otolaryngol Head Neck Surg. 2016154(3):405-20. Essential

research that justifies study of decision making in regional varaitions of tonsillectomy rates. Additionally learnt skills on systematic review.

Abbreviations

A&E	Accident & Emergency
AHP	Analytical Hierarchical Process
CI	Confidence interval
CALIBER	ClinicAl research using LInked Bespoke studies and Electronic health Records
CCG	Clinical Commissioning Group
CPRD	Clinical Practice Research Datalink
DART	Decision-making for Adults with Recurring Tonsillitis study
DoH	Department of Health
EHR	Electronic Health Record
GP	General Practice
HES	Hospital Episode Statistics
HRQoL	Health-related quality of life
IMD	Index of Multiple Deprivation
IRR	Incidence Rate Ratio
ISAC	Independent Scientific Advisory Committee
MINAP	Myocardial Ischaemia National Audit Project
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NNT	Number Needed to Treat
ONS	Office of National Statistics
OR	Odds Ratio

PARTT	Preferences in Adults with Recurring Tonsillitis Tool
RCS	Royal College of Surgeons
RR	Relative Risk
SIGN	Scottish Intercollegiate Guideline Network
SCV	Systematic Component of Variation
UCL	University College London
URTI	Upper respiratory tract infection

Glossary of commonly used terms

A & E: Accident and Emergency department - a medical treatment facility specialising in emergency medicine, the acute care of patients who present without prior appointment; either by their own means or by that of an ambulance. The emergency department is usually found in a hospital.

Baseline biosociodemographics data: For participants of FluWatch a baseline visit was made to the household at enrolment, during which a research nurse assisted families with a series of laptop-based surveys collecting information on basic demographics, health and chronic illness, respiratory hygiene, household structure and relationships, accommodation, contacts and activities. From the third season onwards, these baseline data were self-completed by participants using a bespoke online survey.

BMI: Body Mass Index approximates adiposity by comparing weight to height (weight in kilograms over height in metres squared). However, it does not differentiate lean mass from fat mass.

Completed weeks analysis: Analyses in FluWatch were undertaken in all weeks where there was a completed household weekly status report (see household weekly status report) attached to their household for that week, even if there no individual patient illness data for that week.

CALIBER: Clinical research using bespoke database of electronic patient medical records. CALIBER links the world's largest primary care database (see CPRD below) with secondary care health data (see HES below)

CCG: Clinical Commissioning Groups were originally created to cover a population of around 100,000 UK residents each and were given a yearly budget based on the population size and predicted health care need. The main remit of the CCG was to ensure the most efficient use of their residents' healthcare fund.

CPRD: Clinical Practice Research Datalink – a governmental, not-for-profit research service, that collects primary care medical records data from over 5 million active patient records (and over 13 million overall) drawn from approximately 650 primary care practices in the UK. The diagnosis and management of patients attending their GP is coded using READ codes (see

READ codes) and stored in electronic medical records which are sent to CPRD for anonymization. The primary purpose of coding is for routine clinical use and not research.

Daily Illness report: Participants of FluWatch were asked to complete a daily health diary during days of respiratory illness. The diary requested information on illness onset date, temperature and presence and severity of symptoms such as feeling feverish, headache, muscle aches, cough and sore throat. Diaries also collected data on contact patterns and activities before and during illness, including consultations and antibiotic use.

FluWatch: Cohort study to investigate community burden of influenza and associated medical consultations, run over 6 seasons between 2006-2011

GP: General Practice – a primary health care service located in the community to improve the health of local residents through health promotion, surveillance, screening, deployment of vaccination programmes, disease treatment and medical condition management. There are 7,875 General Practices in England with 57.2 million registered patients.

HES: Hospital Episode Statistics - HES is a data warehouse containing details of all admissions, outpatient appointments and A&E attendances at NHS hospitals in England. This data is collected during a patient's time at hospital and is submitted to allow hospitals to be paid for the care they deliver. HES data is designed to enable secondary use, that is use for non-clinical purposes, of this administrative data. Diagnostic data is coded using the ICD-10 system (ICD-10), whereas procedures are coded using the OPCS4 system (see OPCS4). The CALIBER dataset of combined CPRD-HES data only had access to inpatient HES records from January 1997.

Household weekly status report: The lead member of every household in the FluWatch study was actively contacted every week with automated telephone calls to assess the presence or absence of respiratory illness in each household member.

HPA: Health Protection Agency - a non-departmental public body set up by the UK government in 2003 to protect the public in England from threats to their health from infectious diseases and environmental hazards. In 2013, the HPA became part of Public Health England, a new executive agency of the Department of Health (DoH).

HSB: Help Seeking Behaviour – a series of well-ordered and purposeful cognitive and behavioural steps, each leading to specific types of solutions, depending upon the person's

recognition, insight and dimension of the problem and resources available for problem resolution. In FluWatch we measured HSB by asking patients if they sought help for their symptoms and asking which of the following avenues of help was sought: Pharmacist, GP (see GP), NHS Direct (see NHS Direct), hospital, A & E (see A&E), Other.

ICD-10: 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD), a medical classification list by the World Health Organization (WHO). It contains codes for diseases, signs and symptoms, abnormal findings, complaints, social circumstances, and external causes of injury or diseases. It has been mandated for use in the UK since 1995 and is used for coding health encounters experienced at secondary care level.

IMD: Index of multiple deprivation - is a UK government qualitative study of deprived areas in England. Categorising geographically on areas with similar social characteristics and population of around 1500 people the team calculated an index of deprivation for these geographical units based on income, employment, health deprivation and disability, education skills and training, barriers to housing and services, crime and living Environment. In my analyses, I used quintiles of this index to estimate if the participant lived in the most deprived fifth or the least deprived fifth of the country.

MINAP: Myocardial Ischaemia National Audit Project – a national registry of patients admitted to hospitals in England and Wales with acute coronary syndromes (ACS), established in 1998 to provide participating hospitals with a common mechanism for auditing performance against standards defined in the National Service Framework for Coronary Heart Disease.

Need: In relation to healthcare the term ‘need’ is defined as the population who could benefit from a treatment

NHS Direct: Health advice and information service provided by the National Health Service (NHS), established in March 1998 for residents and visitors in England. The nurse-led telephone information service provided healthcare advice 24 hours a day, every day of the year through telephone contact. It was discontinued on 31 March 2014.

Obesity: Large weight in relation to height. If patient weight to height ratio (see Body Mass Index) is more than 20% of expected then they are defined as obese (specifically defined as BMI≥30 kg/m^2). Obesity is a strong risk factor for many other conditions, including tonsillitis

OPCS4: The procedural classification used by clinical coders within National Health Service (NHS) hospitals based on the earlier Office of Population Censuses and Surveys Classification of Surgical Operations and Procedures (4th revision)

ONS: Office of National Statistics which provides information on deaths and social deprivation

OSA: Obstructive Sleep apnoea, a syndrome defined by obstruction of upper airways during sleep. The blocked airway results in lower blood oxygen levels and greater pressures exerted on the heart. Patients complain of poor sleep and daytime tiredness. Severe OSA can lead to heart failure. Many factors have been shown to be related to the occurrence of this syndrome from obesity (BMI > 30) to large tonsils. Surgical intervention can be advocated for OSA clinically when symptoms are severe and interfere with daytime behaviour, e.g. increased day time tiredness or poor concentration. Tonsillectomy is the most common surgical procedure undertaken for this syndrome in children.

PCT: Primary Care Trust, see CCG

Poor responders: Participants of FluWatch from households that responded to less than 70% of all household weekly status reports (see Household weekly status report) during the period they were enrolled in the study.

Population density: In the FluWatch study the population density of the participants' region of residence was calculated based on their post code as urban for postcodes that mapped to cities or towns or rural for postcodes that mapped to villages or hamlets.

READ code: Standard clinical terminology system used in General Practice in the United Kingdom. It supports detailed clinical encoding of multiple patient phenomena including: occupation; social circumstances; ethnicity and religion; clinical signs, symptoms and observations; laboratory tests and results; diagnoses; diagnostic, therapeutic or surgical procedures performed; and a variety of administrative items. This is the coding system used by practices that are associated with CPRD.

Recurring sore throat phenotype: Any patient who has 3 or more sore throat consultations (see sore throat consultation) within a year, with each consultation being more than 21 days apart from each other (discussed in chapter 2).

Recurring tonsillitis: Defined by SIGN as 7 episodes of tonsillitis within 12 months/5 episodes a year over 24 months/ 3 episodes a year over 36 months. Also see tonsillitis

Sore throat consultations: A collection of READ codes (see READ codes) that describe potential patient consultations in primary care that were related to sore throat infection. Used in CALIBER study to assess the burden of sore throat in primary care.

Sore throat infection: Acute episode of symptoms self-reported by a patient as moderate or severe sore throat for two or more consecutive days, in absence of a cough, with no respiratory tract symptoms in preceding 7 days. Illness was considered resolved when participant was free of symptoms for 2 days or more.

SCV: Systematic Component of Variation measures the degree of variation between regions. The SCV is an adaptation of the proportional hazards model. The method described subtracts the random error (estimated through generalised linear models above) from the estimated total variance to calculate the systematic component of variation. This measure allows comparisons of variability between regions but makes few assumptions about the nature of the variation and allows for appropriate amounts of sampling variation in the data. It has been suggested that variations giving SCVs greater than 3 are likely to be due largely to differences in practice style or medical discretion, and that high variation is described by a SCV of between 5.4 and 10.0, with SCVs greater than 10 being very high variation.

Tonsils: lymphoid tissue situated at the back of the throat and are often the first point of contact our lymphatic system has with bacteria and microbes humans ingest and inhale.

Tonsillectomy: A surgical procedure undertaken under general anaesthesia to remove palatine tonsils. Most commonly undertaken for patients with recurring tonsillitis (see recurring tonsillitis), but also be undertaken for patients with obstructive sleep apnoea (see OSA), or suspected tonsillar cancer.

Tonsillitis: Infectious inflammation of palatine tonsils. Symptoms include severe sore throat, feeling feverish, pain on swallowing food and tenderness in the upper neck. Examination

shows exudate on tonsils, temperature above 38°C and swollen glands in upper neck. Commonly caused by viruses and bacteria. If bacterial, can be complicated by localised abscess, or non-suppurative conditions from acute rheumatic fever to glomerulonephritis. Treated with painkillers and rest. If symptoms severe can be managed with a delayed antibiotic prescription. Can become recurrent in certain groups. Also see recurrent tonsillitis.

UCL: University College London - a public research university in London, England, and a constituent college of the federal University of London. It is the largest postgraduate institution in the UK by enrolment

Contents

Table of Figures

Table of Tables

CHAPTER ONE: Geographical surgical rate variation and the case of tonsillectomy

Chapter Synopsis

In this chapter, I have used historical context to define the importance of regional tonsillectomy rate variation. I have critically evaluated studies, specific to tonsillectomy and then surgery in general, to present two current theories of what drives regional tonsillectomy rate variation. Subsequently I have summarised strategies that have been employed to directly reduce regional tonsillectomy rate variation. Finally, I have outlined evidence from multiple disparate sources (health economics, decision analysis, healthcare equity science) and summarised two preparatory studies I had undertaken to define key reasons why these strategies may have failed. I have concluded this chapter with the aim and objectives of my thesis and an overview of thesis architecture

Introduction to tonsils, tonsillectomy and regional tonsillectomy rate variations

Tonsils are lymphoid tissue situated at the back of the throat (Figure 1 Tonsils) and are often the first point of contact our lymphatic system has with bacteria and microbes that we ingest and inhale. Tonsils produce antibodies that help fight against bacteria and T-cells that attack cells infected with viruses(1).

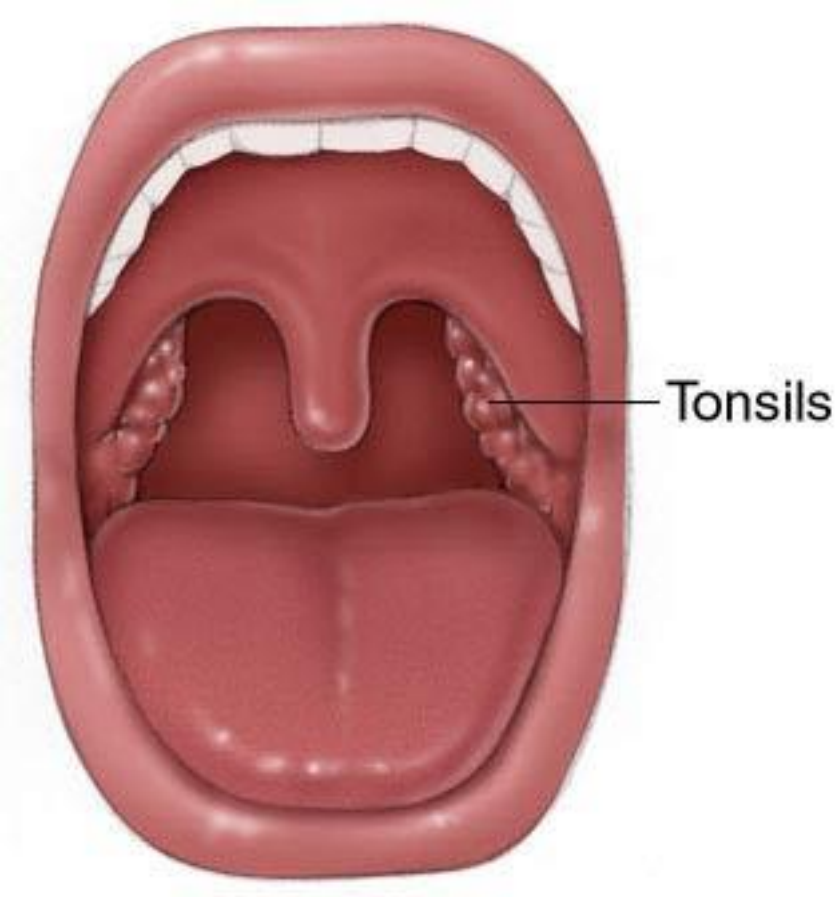


Figure 1 Tonsils.

This diagram shows the location of pharyngeal tonsils, which are lymphoid tissue, ideally located to be the first line of defence against ingested and inhaled pathogens.

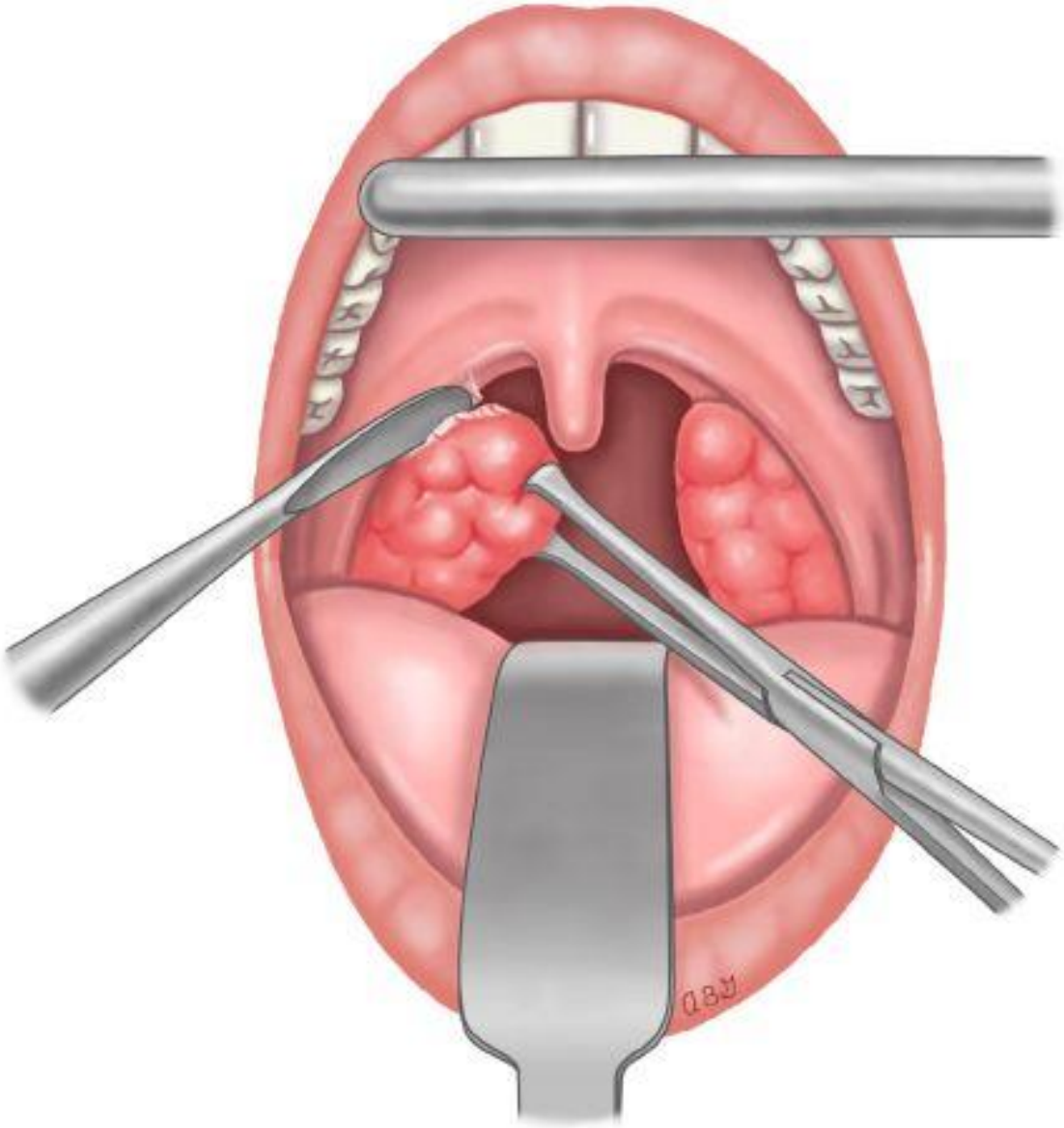


Figure 2 Tonsillectomy.

This diagram shows one method of removing the tonsils surgically – tonsillectomy – which is a procedure undertaken most commonly when patients suffer from recurring infections of the tonsils.

Occasionally the tonsils, and supporting lymphatic tissue, fail to prevent a throat infection (pharyngitis) and can indeed become infected themselves (tonsillitis), causing a pain in the throat that gets worse on swallowing solids and liquids. Sore throats are the second most frequent respiratory infection seen in primary care (2). While in many cases sore throat is

relatively minor and self-limiting, a significant number of patients have recurrent episodes of sore throat with illness, and loss of education or earnings(3-8). In those patients, tonsils can be surgically removed by a procedure called a tonsillectomy Figure 2 Tonsillectomy.

Tonsillectomy is amongst the top 20 most performed operations in England (9) with more than 46,000 undertaken each year. This intervention was most commonly used to treat recurring tonsillitis. However, in recent years, obstructive sleep apnoea may have become a more frequent indication, in children at least (10).

Variation in tonsillectomy rates has been investigated and publicly reported since the 1930's, originally in the UK but subsequently in many other developed countries (11-15). The rapid rise in use of tonsillectomy in the paediatric population following the First World War was probably a key reason as to why so many epidemiological studies used tonsillectomy to investigate geographical surgical rate variation. Epidemiological studies summarised by Glover suggested that one third to one half of all children received this operation before the age of 15 in UK and USA in the 1930s – with nearly 6 times the number of children receiving this operation at its peak compared to modern times (16).

Glover's (16) seminal paper, published in 1938, investigated geographical variation in adenotonsillectomy rates of children 5-14 years old across England and Wales (using school medical records). He found that tonsillectomy rates varied threefold between geographical regions, even when these regions were adjacent. Ever since, regional tonsillectomy rate variation has continued to be topical and controversial.

Bloor (17) undertook an investigation between two socio-demographically similar regions in Scotland known to have wide variation in tonsillectomy rates in children 15 and under. He analysed all referrals received by the ENT departments, between 1961-1970, and showed that referral rates were not statistically different between regions, but tonsillectomy rates were different.

A retrospective observational study of surgical admissions to the Oxford region, UK (18), showed that paediatric tonsillectomy rates varied widely between regions in Oxford. In response, Suleman used NHS Hospital Episode Statistics (HES) data to calculate tonsillectomy rates in patients under 15 years across 380 local authorities over the years 2000-2005(19). In addition, he extracted similar figures from BUPA, the UK's largest private provider covering

40% of England's private medical practice. He found up to 7-fold difference in tonsillectomy rates between regions. There was no correlation between the rates of NHS and private tonsillectomy; a low NHS tonsillectomy rate within a region did not mean that the private rate was greater.

In 2009 the UK Government commissioned the King's fund(20) to undertake an investigation of healthcare variation. They also reported a 7-fold disparity in paediatric tonsillectomy rates between the highest and lowest utilising regions. Compared to other procedures they reported that tonsillectomy remained amongst the five procedures with the highest geographical variation. Not only has high regional tonsillectomy rate variation been noted within other countries (11-15) , but also between countries and types of health care systems (21).

What do we think is driving regional tonsillectomy rate variation?

Historical evidence from survey, epidemiological and qualitative studies has suggested that regional variations in tonsillectomy rates may, in part, be related to "professional uncertainty" amongst surgeons of how to manage the condition which then allows regionally aligned surgeon treatment preference - "the surgical signature" - to drive the treatment decision.

Professional uncertainty in the management of recurring tonsillitis

The first sign that there was professional uncertainty in the management of recurring tonsillitis came from a study of paediatric tonsillectomy in the USA. In 1954, Bakwin (22)randomly sampled 1,000 children, 11 years of age, from the public schools of New York City. He noted that 61 per cent had already received a tonsillectomy. The remaining 39 per cent were examined by a group of surgeons who selected 45 per cent of these for tonsillectomy and selected the remainder for non-surgical treatment. The children that were selected for non-surgical management were then re- examined by another group of surgeons, who recommended for tonsillectomy 46 per cent of those remaining. The authors commented that 935 out of the 1,000 children would have had a tonsillectomy if they had followed this pathway and concluded that there was no clinical consensus amongst surgeons as to who would benefit from the surgery (22).

This was backed up by Bloor's (23) qualitative study of 6 ENT surgeons from a high tonsillectomy-, and 5 from a low tonsillectomy-rate region in Scotland, who undertook 493 sequential consultations with patients (15 years and under) referred with recurrent sore throat. He reported considerable variation in decision-making and concluded variation in tonsillectomy rates was due to the variation in the interpretation of an incomplete evidence base.

The theory of professional uncertainty driving regional tonsillectomy rate variations was further supported from observational studies around other surgical procedures that demonstrated greater variation for procedures for which there was a professional uncertainty on how cases should be managed, compared to similar procedures where management had been standardised (17,24-26). For example, there was low regional variation in paediatric herniorraphy compared to paediatric appendicectomy (24) and low regional variation in hip fracture surgery compared to hip replacement surgery (27). Qualitative and observational studies of surgical management of breast cancer and osteoarthritis, amongst other conditions, led to a better understanding of professional uncertainty (discussed below).

By critically evaluating the broader evidence base around professional uncertainty I have identified three scenarios that have been found associated with professional uncertainty: 1. Conditions that have equipoise across treatments (e.g. management of early breast cancer where modified radical mastectomy has equivocal survival to breast conserving surgery (28-30); 2. Conditions for which there is no consensus amongst surgeons about how patients should be managed, e.g. breast cancer (31,32)) 3. Conditions for which there is insufficient evidence about outcomes, e.g. Meniere's disease (33,34). Conditions that fulfilled any of these criteria have in the past been defined as 'preference-sensitive' since 'best treatment' could only be defined as the treatment that performed best on the outcomes important to that individual patient (24,35). Wennberg suggested that when there was no consensus on how to manage a condition there was greater opportunity to consider non-medical factors, such as surgeon preferences and social norms.

This theory may have a part to play in explaining why certain procedures were more prone to regional surgical rate variation compared to others. It did not, however, explain patterns of variation in surgical practice for the same condition – that is if there was uncertainty in the evidence for a management of a specific condition then all surgeons in all regions would be

exposed to that uncertainty. It would fail to explain why uncertain surgeons in some regions were more prone to surgical intervention than their counterparts in a neighbouring region. A further theory was developed to explain this observation – this theory was based on the idea of surgeon treatment preference and was coined the "Surgical Signature".

Surgeon treatment preference

Evidence for surgeon treatment preference influencing regional surgical rates originally came from a "natural experiment". Whilst Glover (16) was investigating regional paediatric tonsillectomy rates in English boroughs, he noted that Hornsey had a higher rate of tonsillectomy than all neighbouring regions. Half way through his observation window a new school district surgeon was appointed at Hornsey. Within a year, the rates of tonsillectomy in the district dropped by a factor of ten, and remained low for years afterwards. Glover attributed the drop in rates to the change in "surgeon opinion" embodied in the different practice styles of the two different surgeons. He concluded that geographical variations were due to the surgeon's preference towards the operation. In addition, whilst this may have helped explain the regional variation seen in paediatric tonsillectomy rates across school districts that were each administered by a single surgeon, it did not explain the high tonsillectomy rate variation across bigger geographical units that were administered by multiple surgeons. Wright (36) argued that through processes of attraction and retention, practice styles tended to cluster, resulting in geographically based patterns of variation amongst a larger number of surgeons. Chassin (37) suggested that the authoritative nature of continued medical education, which was often regionally based, promoted geographical alignment in surgeon treatment preference. However, the surgical signature hypothesis did not explain the observation that surgeons had previously changed their behaviour, within whole regions, suddenly for non-medical reasons (e.g. changes in the remuneration system) (38).

What has been tried to reduce regional tonsillectomy rate variation?

Whilst tonsillectomy rates were the first surgical procedure found to have such large variation between regions, it soon became evident that this was part of a larger issue (39-41). The issue of regional healthcare variation became popularised in the media as the "post code lottery" (42-44), and was broadly publicised as source of social injustice with certain populations being prevented from having necessary treatments or other populations being

exposed to unnecessary surgery. From a policy perspective, Government bodies saw the same regional surgical rate variation as a source of inefficiency. In fact, a report commissioned by the UK Department of Health in 2009 found that up to £700 million could be saved by de-commissioning operations that had high regional variations (45). NHS England created RightCare in 2009, a national programme devoted to reducing practice variations and securing value for the NHS. It produced atlases of practice variation, including tonsillectomy, and guided local health authorities towards prioritising treatments to prevent under- or overuse. Public health officials started reporting large differences in standardised mortality rates, which were up to three times greater in some regions of England compared to others (46), and they argued that regional surgical rate variations were tied into the bigger issue of healthcare inequity – that is a failure to provide equal healthcare to those who had equal need.

The Evidence Based Medicine movement (EBM) grew in direct response to variation in health care (47). Three major consequences of this movement included building the knowledge base for implementation and knowledge translation through promotion of high level research (48), the establishment of the Cochrane Collaboration(49), which was tasked with the collation and summarising of evidence from clinical trials, building national and international infrastructures for developing and updating clinical practice guidelines(50).

Specific to tonsillectomy, multiple randomised controlled trials were initiated to evaluate the effectiveness of this surgery for the treatment of children (3,51-53) (52) and adults (5,7) with recurring tonsillitis. The Cochrane collaborative recently updated a meta-analysis of the above randomised controlled trials(54) and concluded that whilst in children (adeno)tonsillectomy resulted in a reduction of 0.6 sore throat episodes (95% confidence interval -1 to -0.1) in the first post-operative year, there was insufficient evidence with regards to adults to guide clinical decisions. In parallel to this, the National Institute of Health Research (NIHR) Health Technology Assessment (HTA) programme commissioned a randomised controlled trial assessing the effectiveness of tonsillectomy in adults with recurrent tonsillitis, which is ongoing (55). A better understanding of the risks and complications of tonsillectomy came following on from the National Prospective Tonsillectomy Audit, which was undertaken by the Royal College of Surgeons in 2005. The audit included 33,921 tonsillectomies and found a total primary and secondary bleeding rate

of 3% and 0.9% returned to theatre for arrest of haemorrhage(56). With the results of these papers a better understanding of potential benefits and risks of tonsillectomy was now developed.

Based on the above evidence the Scottish Intercollegiate Guidelines Network (SIGN) generated GP referral advice and ENT criteria for tonsillectomy(57). This guidance was endorsed by ENTUK. Tonsillectomy was advocated on children and adults if:

- sore throats were due to acute tonsillitis AND
- the episodes of sore throat were disabling and prevent normal functioning AND
- 7 or more well documented, clinically significant, adequately treated sore throats in the preceding year OR
- 5 or more such episodes in each of the preceding two years OR
- 3 or more such episodes in each of the preceding three years

Shortly after the release of the guidelines a survey of ENT surgeons showed that whilst many had criticisms, 84% of respondents reported following the guidelines(58).

Yet despite this growing evidence better defining benefits and risks of tonsillectomy, development of national guidance and high compliance reported by ENT surgeons, the King's Fund (20) and the NHS RightCare programme (59) reported in 2011 that regional variation in tonsillectomy rates increased and got worse in their rankings.

What are the potential reasons that we have failed to reduce regional tonsillectomy rate variation?

Following on from critical evaluation of the published work around regional surgical rate variation, I have hypothesised three potential drivers of why regional tonsillectomy variation has failed to reduce despite the above strategies:

1. Regional surgical rate variation is a response to regional need (e.g. rates of recurring tonsillitis are different between regions)

2. Regional surgical rate variation is a component of shared decision making (e.g. regional differences in the way the medical consultation is conducted encourage treatment decisions to align with regional norms)
3. Regional surgical rate variation is a response to regional healthcare infrastructure (e.g. regional healthcare policy encourages differences in tonsillectomy rates between regions)

Regional surgical rate variation is a response to regional need

Regional surgical rate variation may be a marker of a healthcare system that is adaptable to the 'need' of the local populations. The term 'need' is defined as the population who could benefit from a treatment (60) and has historically been divided into three broad categories:

1. The true incidence of surgically treatable disease, which may be related to demographic, environmental and lifestyle factors

2. How frequently sub-clinical disease is detected with medical testing, which may relate to help seeking behaviour such as GP consultation

3. The preference of patients to undergo surgical intervention, which may be related to social norms

Variation related to true incidence of disease

It may seem obvious that as numbers of people who may benefit from surgery increases, so does the rate of surgery. This phenomenon becomes apparent for conditions that are easily diagnosed, always detected and recorded, unlikely to be falsely recorded, and for which the only treatment is surgery (i.e. there is little room for choice or preference). A good example is hip fractures, which all require hospitalisation and there is only one agreed treatment: Open reduction and internal fixation (61). Rates of hip fracture surgery in Hawaii are at least 60% lower than elsewhere in the US because fewer patients have hip fractures (62). For diseases that are not as easily recorded (e.g. ischaemic heart disease), the incidence can be inferred

through risk factors. For example, the North of England has markedly higher rates of all types of cardiovascular interventions than other regions of the country, driven by much higher prevalence of cardiovascular risk factors, such as hypertension (63), diabetes (64,65), and smoking (66), and therefore, higher incidence of ischaemic heart disease. In the case of tonsillitis, it is still unclear how regional rates of tonsillectomy relate to regional rates of tonsillitis.

II. Variation related to differences in disease detection

Even when the true prevalence of disease varies little by geography, the number of surgically treatable patients could vary according to regional differences in diagnostic testing of patients with asymptomatic or subclinical disease. For example, while there is little evidence that the true incidence of prostate cancer varies widely within countries, rates of prostate specific antigen (PSA) screening—the most common means by which this disease is detected—differ markedly. These variations in screening are strongly correlated with variations in prostatic biopsy and resection rates (67). Lu-Yao and colleagues found that PSA screening rates differed five-fold between two North American cities and suggested that this may have contributed to the five-fold difference in prostatectomy rates between the two(68). Specific to tonsillectomy, most tonsillitis cases can be safely managed in the community and so detection requires a patient to present to their General Practitioner (GP) or other health care provider, therefore, detection rates of tonsillitis are highly influenced by consultation rate for sore throat. Studies have shown that GP consultation for acute respiratory infections varies with patient characteristics such as age, sex (69) (70-72) marital status(73), family size (73,74), education level (75)and social class (76), all factors that are also known to vary between regions. In the case of tonsillitis, it is still unclear how regional rates of GP consultation for tonsillitis relate to regional rates of tonsillectomy.

III. Variation related to patient choice

Finally, differences in patients' willingness to undergo surgery may play a part in regional variation in procedure rates. Hawker et al. conducted a study in patients with hip and knee osteoarthritis residing in two areas of Canada, one with low and the other with higher rates of hip and knee arthroplasty. For patients judged clinically appropriate for surgery, the investigators then presented patients with detailed information about the nature of and risks and benefits of joint replacement. Only 8·5% of patients in the low-rate area expressed that

they were "definitely willing" to have surgery, while 14·9% of patients in the high-rate area expressed that preference(77). There is growing evidence from survey studies to suggest that patients in high surgical areas may have a higher preference for surgery than those in low surgery areas (78-80). Other studies suggest that patient preference for surgery may vary by patient characteristics known to vary between regions, for example age and gender (81-84), ethnicity (85-88), English language fluency (89), education level (90) and socioeconomic status (91). However, these studies ask about hypothetical willingness to undergo surgery, which may not be reflective of decisions reached in a real consultation. Additionally, measurement of willingness for treatment based on selecting an ordinal value on a spectrum from "*definitely not willing to have to treatment*", through to "*definitely willing to have treatment"*, is less reliable since it uses a single question to measure a complicated construct. Studies have used willingness-for-surgery as a surrogate for patient-preference-for-surgery since the latter has historically been difficult to elicit in a reliable and efficient manner. In the case of tonsillitis, there is no instrument that measures treatment preference and it is still unclear how regional rates of tonsillectomy relate to regional preferences towards tonsillectomy or non-surgical management of recurring tonsillitis.

"Need" in relation to tonsillectomy

The difficulty of relating regional tonsillectomy rate variation to regional variation in 'need' for tonsillectomy comes from three main issues:

(1) Tonsillitis can be managed conservatively without medical intervention, and so the true incidence of the disease has to be measured in the community, not health care settings, to prevent underestimation of the disease burden. Monitoring short lived infections in the community, and resulting increases in consultation rate for sore throat has been notoriously difficult given the number of people that need to be included in a cohort and followed up intensively for long periods of time to truly estimate the community incidence of a condition.

(2) Diagnosis and medical management of tonsillitis is most frequently undertaken and therefore recorded in primary care, while treatment of recurring tonsillitis (tonsillectomy) takes place and is recorded in secondary care. There is no official national linkage between primary and secondary healthcare datasets. Whilst patients in the UK have a unique health

code that identifies them in both primary and secondary care, government attempts to link and centralise health records across all settings have not been successful.

(3) There is more than one way to treat recurring tonsillitis and so variations may reflect patient choice – which has historically been very difficult to measure. Measures such as visual analogue scores have issues with reliability as they use a single question to approximate a complicate concept, whilst using interview techniques based on standard gamble or time-trade off are labour and time intensive.

In part because "need" for tonsillectomy is so difficult to ascertain, previous studies of regional tonsillectomy rate variation have just assumed that the "need" for tonsillectomy does not vary between regions.

Regional surgical rate variation is a component of shared decision making

The surgical consultation is the key setting where decisions are finalised about a patient having surgery or not. There is good qualitative evidence suggesting that aspects of this decision-making process varies with factors that are known to vary between regions. For example, the amount of information given during a consultation varied according to patient characteristics such as age (92), gender (93), social class (94) (95) , educational status (96) and ethnicity (97), as well as surgeon characteristics such as income, social class background, political ideology (98) and race (99). Similar evidence exists regarding type and detail of information. This was partly thought to relate to the doctor's stereotyped perception of patients' needs (99) and partly to do with information seeking behaviour that was more common to certain patient groups (94) (97) (93). Whilst these early studies suggested that variations in population and surgeon characteristics may influence the shared decision making process there was no previous research that directly examined the role of the surgical decision-making process (or components therein) on treatments chosen and therefore, regional surgical rate variation.

Studies of patient decision aids may have indirectly investigated this association between the medical decision-making process and regional surgical rate variation. Patient decision aids are tools that help patients make difficult decisions based on the most recent medical evidence and patients' personal treatment preferences.

In a collaboration with researchers at Johns Hopkins Medical School, I undertook a systematic review of studies reporting on the use surgery-related patient decision aids and its effect on the treatment choice (100). We found 17 studies that were relevant to our research question. Most studies (n=10) showed that patient decision aids influenced patients' treatment preference. Nine found that exposure to a decision aid was likely to change patient treatment decisions away from surgery (e.g. 26 – 38 % reduction of patients choosing joint-replacement surgery), whilst one showed an increase in numbers of patients choosing surgery (change in those choosing laminectomy to treat lumbar disc herniation went from 26.7% to 35.8%). Whilst our review suggests that factors within the consultation may have a role in treatment decision, and therefore, surgical rate variations, there was insufficient evidence to understand which part of shared treatment decision was most related to the change in regional surgical rates.

This has been difficult to investigate because as in other conditions, the surgical consultation is a complicated shared decision making process that contains multiple interconnected concepts. There is little understanding of which concept could lead treatment decisions to line up with regional norms, and even less understanding of how accurately these concepts can be measured.

Regional surgical rate variation is a response to regional healthcare infrastructure (supply sensitive variations)

Supply-sensitive variations have been traditionally described in relation to chronic conditions where the use of service (e.g. hospitalisations, diagnostic tests, Intensive Care utilisation) varies in direct proportion to the capacity of the system (number of acute beds, number of physicians, number of Intensive Care beds etc.) (101). Whilst supply sensitive variations could result in regional variations in elective surgery, there has been no published evidence to suggest such an association.

Measuring regional differences in capacity for tonsillectomy is difficult as it may relate to the number of beds, numbers of other ENT surgeries being undertaken by a department, number and type of surgeries being undertaken by other surgical specialties reducing access to

theatre space, number of clinics running, number of specialists and local policies that promote or prevent access for tonsillectomy. Whilst his has never been measured, there is some evidence using the number of ENT surgeons per capita as a surrogate marker for capacity that suggests capacity does not vary between regions. Specifically, there does not appear to be a correlation between the number of ENT specialists within each country and number of paediatric tonsillectomies between those countries (e.g. USA has the same number of ENT specialists per 100,000 population but has less than half the rate of tonsillectomies compared to Netherlands)(21).

Whilst the number of specialists available may not drive regional variations in tonsillectomy rates, there may be a role for regional health care policies.

Within the National Health Service, the Government funds health services through regionally based local health authorities – called Clinical Commissioning Groups (CCGs). The CCGs were originally created to cover a population of around 100,000 UK residents each and were given a yearly budget based on the population size and predicted health care need. The main remit of the CCG was to ensure the most efficient use of their residents' healthcare fund. To facilitate this, the Government provided guidance as to how they could prioritise, or 'ration', treatments to allow most efficient use of their funds for their local populations' needs – called commissioning guidance. Therefore, review of the funding infrastructure within the NHS suggested a potential role for regionally based "rationing" of services (commissioning guidance) in the regional tonsillectomy rate variation. As there was no published evidence on the relationship between local rationing (commissioning guidance) and local tonsillectomy rates, and to better understand regional rationing policies, I sent out a Freedom of Information request to all 211 CCGs. I asked if they had any published commissioning guidance in relation to tonsillectomy, and if they had, what were its contents and when it was placed into action. I received responses from 189 CCGs (89% response rate).

All CCGs had commissioning guidance for whom should receive a tonsillectomy – that is rationing tonsillectomy to those who would have maximal benefit. However, review of the actual guidance documents showed that different CCGs had used different criteria to define who would gain maximal benefit. For example, some regions (11%) required patients to have attended their GP for seven episodes of tonsillitis in the preceding 12 months, whilst others (89%) just required patients to confirm themselves that they had suffered seven episodes of

tonsillitis in 12 months. Some requested that a microbiological swab prove that the tonsillitis was caused by Streptococcus (full responses through correspondence). This meant some regional systems allowed access to tonsillectomy more than others. However, there was no data to show whether the local commissioning guidance were related the rates of tonsillectomy.

Rationale for this thesis

Regional surgical rate variations are important to address for Public Health, as our current healthcare system may be exposing certain groups of our population to unnecessary surgery, whilst preventing others from receiving necessary treatment. I have shown above that tonsillectomy is not only one of the most common operations undertaken in this country, it also has one of the highest rates of regional variation, despite national strategies undertaken to reduce this. Developing a better understanding of the drivers of regional tonsillectomy rate variation therefore may help guide future health policy in the broader context of regional surgical variation. Whilst I have shown there are many aspects of regional tonsillectomy rate variation that remain poorly defined I have focused my thesis on areas where there is preliminary evidence to suggest a link to regional variation, to provide a better understanding of this finding. The need for tonsillectomy is determined in part by recurring tonsillitis and in community sore throat incidence. The review of regional tonsillectomy-rate-variation research shows that there has never been a specific investigation into the regional variation in proxy measures of 'need' for tonsillectomy. There has only been an assumption that 'need' must be same from region to region. And whilst decisions that lead to tonsillectomy may start from the patient seeking help for their first episode of tonsillitis, through to decisions made by the GP to refer the patient whose tonsillitis becomes recurrent to a specialist (or not), there is considerable literature implying that regional tonsillectomy rate variation is the result of region specific surgeon practice styles (the 'surgical signature' hypothesis). Despite this body of evidence there has never been a study that investigates the ENT specialist-patient medical encounter to see how it affects the treatment decisions of patients with recurring tonsillitis and refute or further support this hypothesis.

Aim

The aim of my thesis was to develop a better understanding of the drivers of regional tonsillectomy rate variation by quantifying regional variation of tonsillectomy rates in relation

to regional demands, and by exploring the role of professional uncertainty and treatment preference on the treatment chosen.

Objectives

The objectives were to establish the:

A. Rate and regional variation of self-reported sore throat and in the rate of relevant consultations for sore throat in the community;
B. Rate and regional variation of recurring sore throat in primary care;
C. Rate and regional variation of tonsillectomy in secondary care, after adjusting for local rates of recurring sore throat;
D. Constructs of clinical decision making and thereby ascertain which concepts were most likely to be related to surgical rate variation;
E. Role of surgeon and patient decisional uncertainty on the treatment chosen for recurring tonsillitis;
F. Role of surgeon and patient treatment preference on the treatment chosen for recurring tonsillitis;

Thesis architecture

To address these objectives, I have separated my thesis into two sections:

Epidemiological studies quantifying regional variation in tonsillectomy rates in relation to regional 'need'.

Decision-making study investigating the potential role of professional uncertainty and treatment preference on the treatment chosen. Please see Figure 3 Thesis architecture

Part I is separated into three chapters (Chapters 2-4).

In Chapter 2, I have summarised the cohorts used in the upcoming studies (FluWatch and CALIBER) and why I used them, how I defined outcomes, co-variates, and potential sources of bias.

In Chapter 3, I have reported a population study analysing self-reported moderate-severe sore throat illnesses in the community, and resulting help-seeking behaviour. This chapter

investigated regional variation in sore throat incidence and in the rate of relevant consultations for sore throat after accounting for local population risk factors (Objective A).

In Chapter 4, I have reported a study that used linked primary and secondary care healthcare records to investigate regional variations in recurring sore throat (Objective B) and tonsillectomy rates (Objective C) after accounting for local population risk factors, as a proxy for demand.

Part II is separated in three chapters (chapters 5-7).

In Chapter 5, I have evaluated the main concepts in clinical decision-making, and the instruments available to measure these concepts, with a particular focus on those that may be involved in surgical rate variation (Objective D).

In Chapter 6, I have summarised the development of a novel instrument to measure potential treatment preferences for adults with recurring tonsillitis (Objective F).

In Chapter 7, I have reported the decision-making study investigating the role of decisional uncertainty and treatment preferences in relation to the treatment chosen for adults presenting to ENT clinics with recurring tonsillitis (Objectives E & F).

In Chapter 8, I have brought together my results from both parts together in a discussion of what this thesis adds and what further work is needed.

Figure 3 Thesis architecture

CHAPTER TWO: Creating electronic cohorts from the FluWatch study and CALIBER database

CHAPTER Synopsis

In this chapter, I have summarised the FLUWATCH and CALIBER programmes and the preparatory work I did in this thesis. I have outlined and highlighted key components of data management that were necessary to align existing datasets towards answering the aims of my thesis. I phenotyped dependent and independent variables to ensure I was measuring valid risk factors and outcomes. I evaluated the impact of poor data quality and summarised strategies to deal with the resulting consequences.

Chapter Introduction

This chapter discusses two existing data sources: The FluWatch study cohort and the Clinical research using LInked Bespoke studies and Electronic Health Resources (CALIBER) database of electronic patient medical records. Since the FluWatch study was a population-based study of upper respiratory illness, its resulting database provided an ideal source of information of sore throat infections in the community and associated rate of sore throat consultations, allowing me to address *objective A (Rate and regional variation of self-reported sore throat and help seeking behaviour in the community)*. CALIBER linked the largest primary care database in the world (Clinical Practice Research Database - CPRD) with secondary care hospital data (Hospital Episode Statistics - HES) to provide an ideal source of information on recurring sore throat management in primary care and secondary care, allowing me to address *objectives B (Rate and regional variation of recurring sore throat in primary care) and C (Rate and regional variation of tonsillectomy in secondary care, after adjusting for local rates of recurring sore throat).*

FLUWATCH: Introduction

FluWatch was a collaboration between epidemiologists at the Centre for Infectious Disease Epidemiology at University College London (UCL), virologists and mathematical modellers from the Health Protection Agency (HPA, now Public Health England), immunologists at the Medical Research Council (MRC) Human Immunology Unit at Oxford University and the MRC General Practice Research Framework (GPRF). It was created to estimate community burden of influenza and influenza-like illnesses, generate up-to-date knowledge of demographic, social and behavioural factors affecting influenza transmission, measure antibody and T cell immune responses to influenza and to use knowledge generated to inform modelling parameters. Funded by the MRC it began recruitment in 2006, however, when the H1N1 pandemic arose in 2009 further funding was secured jointly from the MRC and Wellcome Trust, allowing continued follow-up and an expansion in cohort size. Additional study aims were to inform the national and international response to the current and future pandemics. Specific objectives were to examine clinical profiles of illness, estimate population infection denominators, monitor changes in population behaviour, and investigate access to services.

Although focused on influenza, the study collected data on all respiratory illnesses experienced by cohort members, making it ideal for the study of the community burden of sore throat infections and related consultation behaviour.

FLUWATCH: Cohort

Households were recruited from registers of 146 volunteer general practices (GP) across England, who formed part of the MRC GPRF or (from the 2009 pandemic onwards) the Primary Care Research Network (PCRN). Participants were selected from GP lists by computer-based random number generation. GPs sent invitation letters inviting the randomly selected person and their household to participate. This meant that larger households, such as those with children were more likely to be enrolled.

To be eligible to participate, the whole household had to agree to take part in follow-up over the coming winter, with adults aged ≥ 16 years agreeing to have blood samples taken. Exclusion criteria included household size > 6 people, individuals with terminal illness, severe mental illness or incapacity and heavy involvement in other ongoing research. GPs reviewed invitation lists and removed anyone meeting these criteria, before sending letters. Cohorts were recruited to allow follow-up of participants over six influenza seasons—the 2006/07, 2007/08 and 2008/09 periods of seasonal influenza circulation, the summer and winter waves of the 2009 pandemic and the first post-pandemic season 2010/11. From season 3 (2008/09) onwards, previous participants were invited to take part again.

In season 1, invitation letters were sent to 2300 households from 42 practices, and 602 individuals from 243 households agreed to participate. In subsequent seasons the response rate was not monitored as practices (rather than the university study team) sent the invitation letters and not all returned data on the number of invitations sent (

	Nov 2006 to Mar 2007 Season 1	Nov 2007 to Mar 2008 Season 2	Nov 2008 to Mar 2009 Season 3	May 2009 to Sep 2009 Season 4	Oct 2009 to Feb 2010 Season 5	Nov 2010 to Mar 2011 Season 6

GP practices/households/persons (*n*)	42/243/602	43/310/779	37/309/729	41/332/797	127/1460/3552	51/361/901

Table 1). Compared with the English population, young adults, non-White ethnic groups, people living in socially deprived areas and those living in the North of England, West Midlands and London were under-represented in the FluWatch cohort (102,103) (for full breakdown of FluWatch Cohort please see Appendix A Table 35)

	Nov 2006 to Mar 2007 Season 1	Nov 2007 to Mar 2008 Season 2	Nov 2008 to Mar 2009 Season 3	May 2009 to Sep 2009 Season 4	Oct 2009 to Feb 2010 Season 5	Nov 2010 to Mar 2011 Season 6
GP practices/households/persons (*n*)	42/243/602	43/310/779	37/309/729	41/332/797	127/1460/3552	51/361/901

Table 1 Numbers recruited by season of FluWatch Cohort.

This table shows the numbers of GP practices used, households recruited and patients recruited through the 6 waves of recruitment, n=7,360.

FLUWATCH: Data collected

Baseline data

A baseline visit was made to the household at enrolment, during which a research nurse assisted families with a series of laptop-based surveys collecting information on basic demographics, health and chronic illness, respiratory hygiene, household structure and relationships, accommodation, contacts and activities – **baseline biosociodemographics data**. From the third season onwards, these baseline data were self-completed by participants using a bespoke online survey. Overall, 7360 patients were initially consented to participate in the study over 6 influenza seasons, and completed base-line data.

Study data

Once participants had enrolled they were actively contacted every week with automated telephone calls to assess the presence or absence of respiratory illness in each household member – **household weekly status report**. For each respiratory illness, participants were reminded to fill in a prospective paper illness diary – **daily illness report**. The diary requested information on illness onset date, temperature and presence and severity of symptoms such as feeling feverish, headache, muscle aches, cough and sore throat. Diaries also collected data on contact patterns and activities before and during illness, including help seeking behaviour and antibiotic use. This information was completed by proxy for children.

From the third season onwards (November 2008-March 2011), FluWatch household weekly status report was extracted through emails and SMS that directed participants to a custom built website for survey completion, if anyone in their household had suffered an illness in the preceding week. Participants were provided with laminated wipe-clean charts at home to record daily illness reports as a memory aid for when they completed the online survey. In the final season the EQ5D3L generic health related quality of life measure was added to the questionnaire and was completed during every illness. 605 participants were lost to follow before their first household weekly status report. Therefore, there were 6755 participants for which there was illness data, over six influenza seasons. Please refer to Appendix B – Flu Watch data management – for a detailed explanation of how 2 data files were merged to create a platform file for analysis.

FLUWATCH: Data quality

Since I was planning to undertake research on a dataset created from daily self-reported symptoms from members of the public there were three key issues that needed to be considered with regards to the quality of the data: Representativeness, validity and completeness.

The FluWatch dataset has pseudo-anonymised identifiers and did not identify the patients' location except at a crude level (ten regions of England).

Representativeness

The FluWatch study was limited by the difficulty in obtaining a fully representative sample because, although selection was random, acceptance rates were low. To overcome known issues of non-representativeness, analyses reported in my thesis were weighted using ONS census data to ensure results represented the age and regional structure of the country{Statistics:2004vm}. I did not weight on ethnic origin or social deprivation because zero numbers in some groups would have led to instability of weighted measures.

Additionally, the method of recruitment meant that larger households (up to a limit of 6 people) were more likely to be recruited. To overcome this, analyses in my thesis, were weighted to the inverse of family size. Weighting was undertaken in line with previous work done on this cohort(102-104).

Validity

Understanding valid relationships between risk factors and outcomes requires an appreciation of the potential nature of the association, based on the available evidence. To improve the validity of this study considerable effort was placed on defining outcomes, categorising risk factors based on the available evidence and developing a conceptual model of potential associations.

Defining outcomes

Sore throat and all associated symptoms (described in the above section) were reported as absent, mild or moderate-severe.

A sore throat infection

Since my objective in this study was to investigate regional variations in tonsillectomy rates in relation regional variations in the underlying disease burden for tonsillectomy, the ideal measure would have been a community measure of bacterial tonsillitis. However, there remains considerable heterogeneity between clinical signs and symptoms of all acute sore throat infections with no objective measure of bacterial tonsillitis currently available. There is currently no available data on the relationship between acute sore throat infections and tonsillitis. In fact, there is disagreement even amongst GPs following history and physical exam on the diagnosis and coding of tonsillitis (105). Therefore, since I had no way to accurately and reliably measure tonsillitis in the community I chose to use symptoms that most closely approximated bacterial tonsillitis from the available evidence – that is moderate-severe sore throat on 2 or more consecutive days, associated fever, and absence of a cough. This approximated CENTOR criteria for the diagnosis of bacterial tonsillitis(106), but did not include white spots on the tonsils or tender neck lymph nodes – neither of which had been measured in the original Flu Watch study. An episode of sore throat infection was assumed to have ended safely when the participant was free from symptoms for two days or more.

A new episode was recorded after at least seven days without symptoms. A sensitivity analysis was conducted extending this to 14 and 21 symptom free days (see Table 2 Sensitivity analysis when using differing definitions of new sore throat illness.) Since there was little difference in the incidence of sore throat, irrespective of the disease-free interval used, I chose 7 days to maximise sore throat illnesses that could be included.

	Number of cases	Days at risk	Incidence rate/1000person days	P value
Each episode >7 days apart	3337	735315	4.22(3.89-4.62)	<0.001
Each episode >14 days apart	3230	735315	4.11(3.77-4.49)	<0.001
Each episode >21 days apart	3135	735315	4.00(3.66-4.34)	<0.001

Table 2 Sensitivity analysis when using differing definitions of new sore throat illness.

This table demonstrates the number of sore throat illnesses captured when using varying definitions of a new sore throat – either 7, 14 or 21 days after the previous episode finished.

Categorising risk factors

Age was categorised into preschool (0-4 years), school age (5-15) adolescence and young adult (16-24), early adulthood (25-44), middle age (45-64) and retirement age (>65) as upper respiratory tract infections have been shown to be different in these groups (2,107,108). **Ethnicity** was defined as white British and other (small numbers of non-white ethnicities precluded further meaningful sub classification). Postcode was used to define three different variables: 1. Participants' **geographical region** in England (North, West Midlands, East Midlands and East of England, London, South East and South West); 2. **Population density** (defined as urban for postcodes that mapped to cities or towns or rural for postcodes that mapped to villages or hamlets) and 3. **Index of multiple deprivation** (categorised into national quintiles: IMD1 describing the most deprived quintile and IMD5 describing the least deprived quintile). Patients were defined as **vaccinated** if they had received the influenza vaccination specific to the current influenza season. **Health utility/status** was measured using the EQ5D-3L questionnaire, which consisted of five dimensions (**mobility, self-care, usual activities, pain/discomfort and anxiety/depression**), each with 3 levels of functioning (no problems, some problems and extreme problems). Both the index score and domain specific values were evaluated.

Conceptual models

Conceptual models were created to help guide analyses as well as inform the discussion. Therefore, two models were developed to help 1. Identify causes of sore throat in the community (Figure 4 Conceptual model of sore throat infections.) and 2. Identify causes of GP consultation in those with sore throats (Figure 5 Conceptual model GP consultations for sore throat infections.).

1. It was considered that sore throat could be due to infectious or non-infectious causes (e.g. acid reflux). It was considered that infectious causes required contact with the infection and would therefore likely be mediated by contact patterns. The effect of whether a contact leads

to infection could be mediated by hygiene and by immunity. I note that these mechanisms (contact patterns, hygiene, immunity and non-infectious causes of sore throat) were not directly measured. All other risk factors were hypothesised to work through these mechanisms. Age, gender and social deprivation were assumed to potentially affect all mechanisms. Chronic illness was assumed to act through all mechanisms apart from hygiene. Smoking was assumed to act through immunity and non-infectious causes. Urban/Rural status and the number of people in the household were assumed to largely act through contact patterns.

2. The likelihood of consultation was assumed to be potentially mediated by health service accessibility, characteristics of the acute illness, underlying vulnerabilities that may increase overall level of health concern, personalities and social norms. Except for characteristics of the acute illness I did not have direct measurements for any of these mechanisms. Age, gender and socioeconomic status were assumed to potentially affect all mechanisms. Chronic illness was assumed to act through influencing course of disease and overall level of health concern. Population density was assumed to act through health service accessibility. Ethnicity was assumed to act through cultural norms. The characteristics of pain, loss of usual activity, loss of ability to self-care, reduced ability to go out, ear pain and duration of symptoms, were thought to act by being abnormal symptoms.

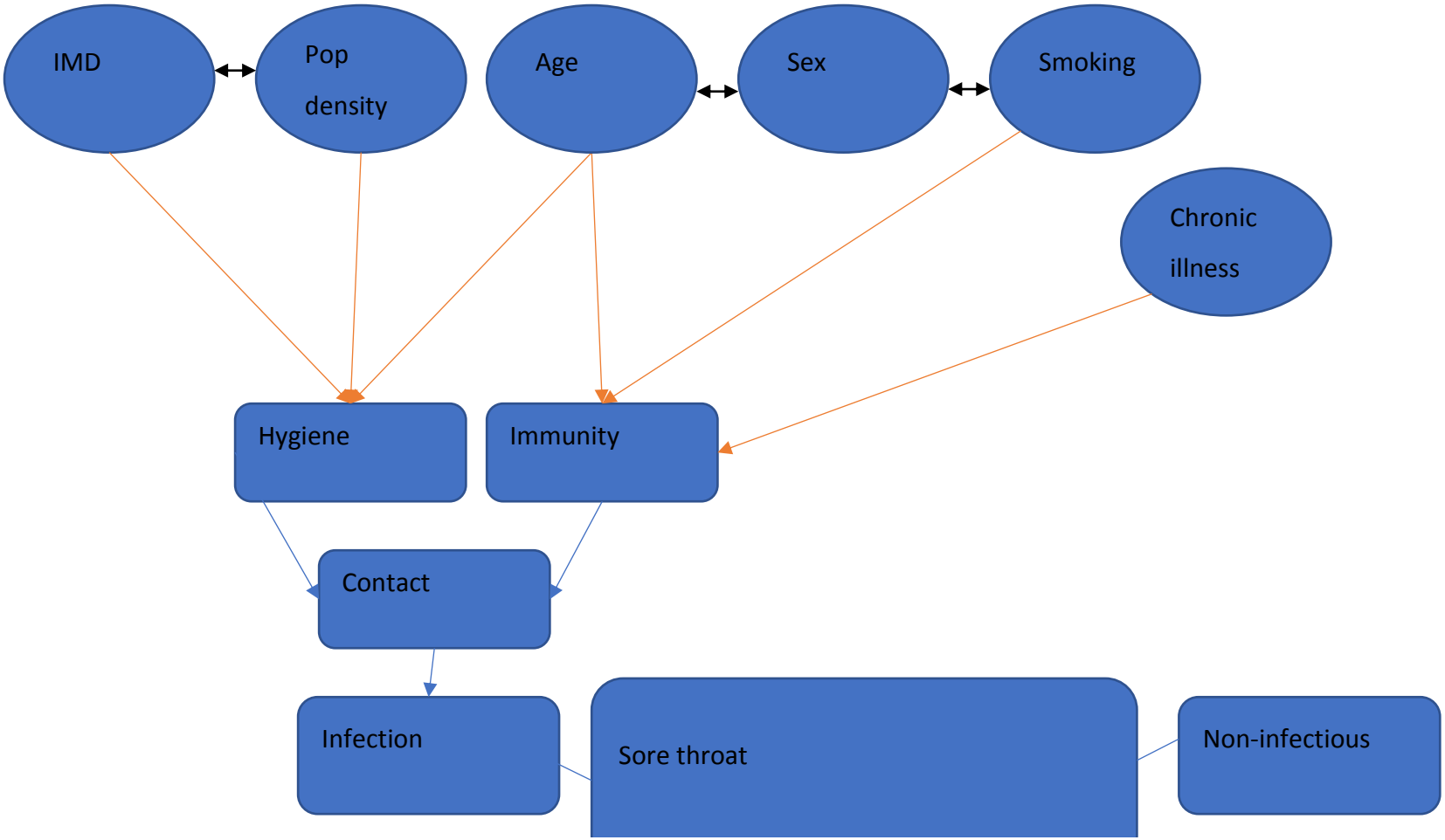


Figure 4 Conceptual model of sore throat infections.

This is a conceptual model of factors that could lead to sore throat infections. Using the infectious disease model, sore throat infections were thought to be propagated by exposure and breakdown in immunity. Available co-variates were considered to act through one or both of these mechanisms

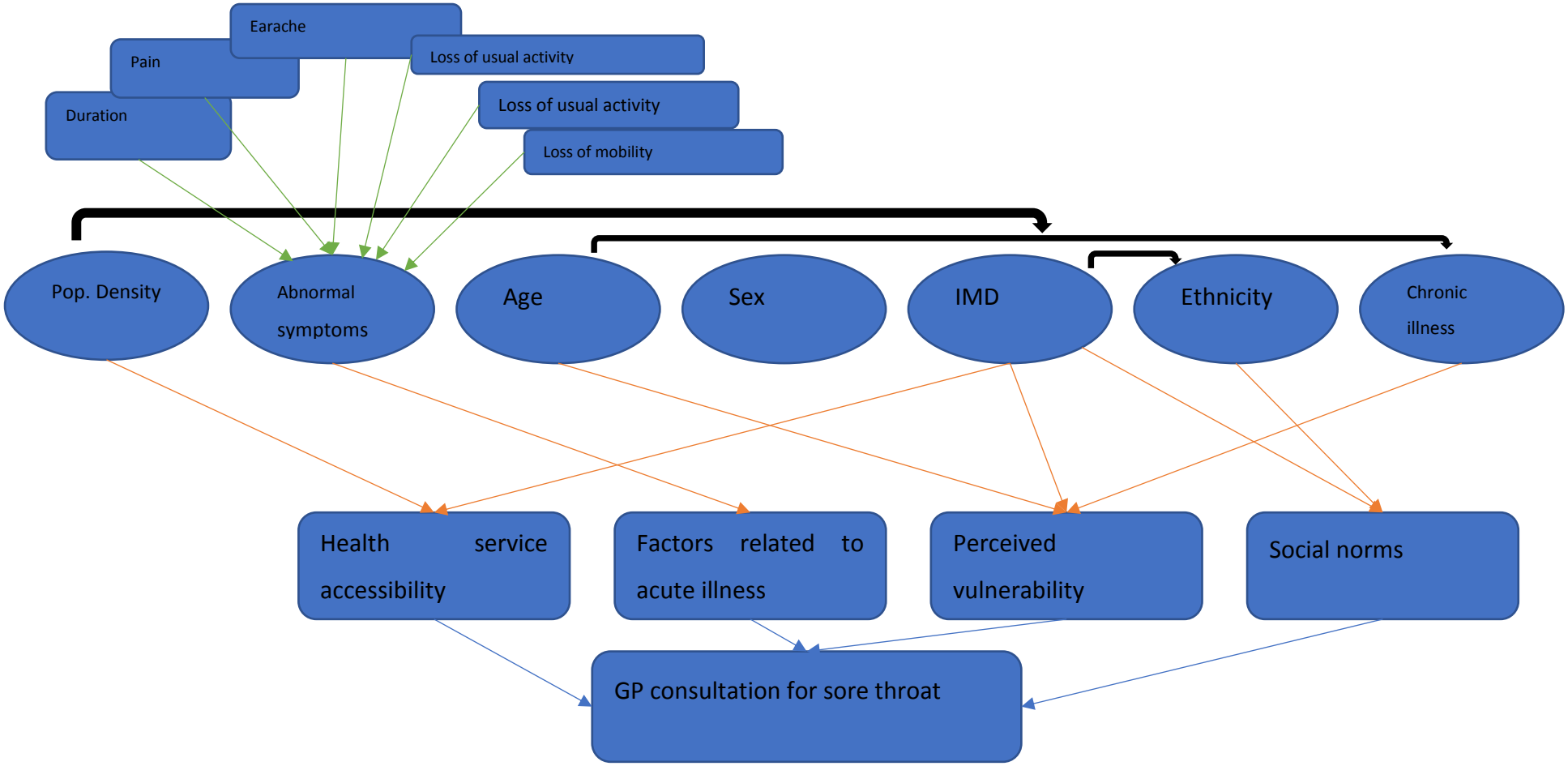


Figure 5 Conceptual model GP consultations for sore throat infections.

This is a conceptual model of factors that could lead to GP consultation for sore throat. The likelihood of a patient seeking a consultation was thought to be dependent on underlying vulnerabilities, social norms, illness severity and accessibility.

Completeness and missing data

Household weekly status report responses were obtained from 88.4% in seasons 1 -2, which increased after the introduction of email and online surveys in season 3 to more than 92%. **Poor responders** were classified as participants from households that responded to less than 70% of all household weekly status reports during the period they were enrolled in the study. Using these criteria, only 12.4% of households were classified as poor responders. Patients were considered **lost to follow up** if there was baseline data but no weekly status reports or the ultimate four weeks (or more) of weekly status reports were missing: 8.2% (605 participants) were considered lost to follow up.

Sensitivity analyses (see Table 3 Sensitivity analysis of missing data.) were undertaken to explore the impact of missing data. In one analysis, I assumed that weeks with a missing household status weekly report were weeks of no illness (*assumed disease absence analysis*), whilst in another analysis I excluded these weeks from analysis (*completed weeks analysis*). *Completed weeks analysis* showed a sore throat rate of 3.99(3.77-4.23)/1000-person days whilst analysis assuming that weeks with missing household status weekly reports were weeks with no illness (*assumed disease absence analysis*) showed a sore throat incidence of 3.45(3.26-3.65)/1000 person days. Restricting analysis to good responders, by removing poor responders from the analysis, showed a sore throat incidence of 3.88(3.50-4.29)/1000 person days.

	Number of cases	Days at risk	Incidence rate/1000person days	P value
Non-reported weeks excluded	3337	735315	3.99(3.61-4.41)	<0.001
Including all weeks in study	3337	818445	3.45(3.12-3.81)	<0.001
Good reporters only	2600	587548	3.88(3.50-4.29)	<0.001
Poor reporters only	737	230897	2.43(2.01-2.85)	<0.001

Table 3 Sensitivity analysis of missing data.

This table shows a sensitivity analysis of sore throat infections depending on whether missing weeks were excluded, considered to have no sore throats, but also compares those who regularly reported illnesses to those who didn't.

Univariable logistic analyses were undertaken to assess which characteristics were associated with being from a household categorised as poor responders (Table 4 Characteristics of poor responders.). Non-whites, smokers, living in regions greater deprivation, living in rural regions, being young, having chronic illness and not being vaccinated were factors that were more likely to be associated with poor responders.

	Good responders	Poor responders	Odds Ratio	P>Z
Males	4177	1259	1.01(0.93-1.10)	0.772
Females	4638	1416	1	
Non white	2067	125	0.67(0.54-0.84)	<0.001
White	7758	317	1	
Smoker	467	202	0.70(0.59-0.83)	<0.001
Non-Smoker	6789	2057	1	
IMD Quin 1	150	264	0.41(0.33-0.51)	<0.001
IMD Quin 2	360	799	0.52(0.45-0.60)	
IMD Quin 3	711	2373	0.78(0.69-0.88)	
IMD Quin 4	722	2450	0.79(0.70-0.89)	
IMD Quin 5	634	2713	1	
Urban	1717	860	1.36(1.24-1.49)	<0.001
Rural	5114	3485	1	
Age				
0-4 years	533	228	0.64(0.53-0.77)	<0.001
5-13 years	1077	363	0.81(0.69-0.95)	
14-24 years	663	256	0.71(0.59-0.84)	
25-44 years	1827	649	0.77(0.67-0.88)	
44-64 years	3103	738	1.15(1.01-1.31)	
>65 years	1612	441	1	
Chronic illness	1363	480	0.83(0.74-0.93)	0.001
No medical problems	7010	2039	1	
Vaccinated	2152	575	1.19(1.08-1.33)	
Not vaccinated	6390	2039	1	0.001

Table 4 Characteristics of poor responders.

This table shows risk factors associated with being a good responder, n = 11,490.

Completed weeks (*completed weeks analysis)* of all responders was chosen as the data from which further analyses were undertaken as it provided greater power, whilst giving no difference in sore throat rates. It was also free from the assumption that weeks with missing household status weekly reports were disease free weeks (*assumed disease absence analysis*). I therefore, report univariable and multivariable analysis based on the completed weeks of all responders. In multivariable analyses, the role of being defined as a good or poor responder was assessed for its impact in the prediction of sore throat.

Summary

In summary, I used the FluWatch database to study community incidence of sore throat. FluWatch is the largest UK based population study using patient self-report of upper respiratory infections to date. Any issues with representativeness of the cohort, were addressed by weighting analyses. Sore throat illness was defined as moderate to severe sore throat reported on 2 or more consecutive days, in the absence of a cough, with a fever reported on any day; recurrence of sore throat illness as sore throat illness with no symptoms in preceding seven days. Considering this was a long-term study where patients were asked to complete daily health diary cards during periods of illness, completeness of the data set was good with more than 80% good responders (>70% response rate to weekly illness reports). I decided to control for whether someone was a good or poor responder rather than exclude them.

CALIBER: Introduction

CALIBER (Clinical research using LInked Bespoke studies and Electronic Health Resources) is a database of linked routinely collected electronic health records (EHR) from England(109), comprising data from primary care (Clinical Practice Research Datalink, CPRD)(109), hospital admissions (Hospital Episode Statistics, HES)(110), the Myocardial Ischaemia National Audit Project (MINAP)(111) and the national death registry at the Office for National Statistics (ONS).

In addition, CALIBER contains small-area indices of deprivation from ONS (Index Of Multiple Deprivation, IMD) linked by the patient's postcode(112). The IMD is a score calculated for each patient's neighbourhood based on social indices such as income, education, and employment.

The data sources complement each other in providing different types of information about a patient's medical history longitudinally, illustrated with a cardiovascular disease example in Figure 6 Data linkage principles in CALIBER.

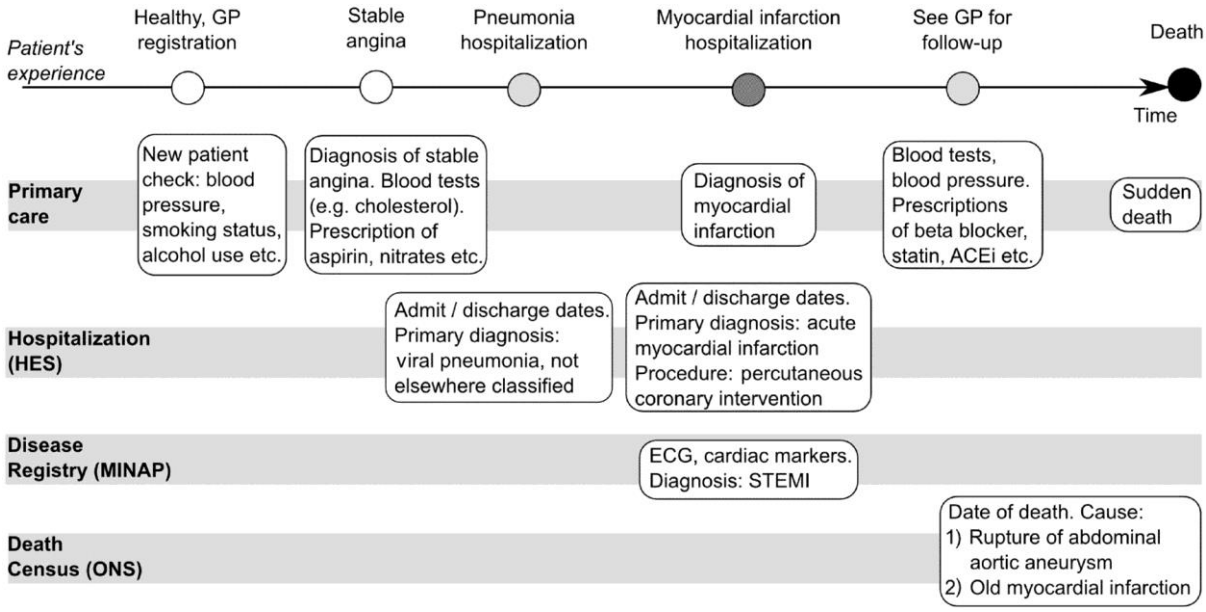


Figure 6 Data linkage principles in CALIBER.

This diagram uses coronary syndromes to show how linkage of multiple datasets could be used to map healthcare journey.

The CALIBER dataset has GP records from 1988 to 2010, but does not include routine information on attendance in hospital outpatients or accident and emergency departments.

The CALIBER dataset has pseudo-anonymised identifiers and does not identify a patient's or general practice's location except at a very crude level (one of 10 regions in England). Further descriptions of the datasets utilised in the subsequent chapters; CPRD, HES and ONS are detailed below.

Access to data

Access to CALIBER data operates by a 'safe haven' model with datasets stored and analysed securely and only aggregate data exportable(113). CALIBER researchers are provided with pseudonymised data (*i.e.,* identifiers such as date of birth, name and address removed) and must commit to not disclosing any information that may be able to identify a patient. Whilst the free text associated with coded data is not routinely available to researchers, it can be requested (with a cost for manual anonymisation) and has been used for validation studies(114).

Ethical and scientific approval

CPRD has Multi-centre Research Ethics Committee approval for all purely observational research using its linked EHRs (CPRD, HES, ONS)(115).

The CALIBER dataset comprises CPRD data linked to HES, ONS and MINAP by a trusted third party with the final dataset held in a pseudonymised form. In addition, the CALIBER record linkage study has had separate ethical approval (09/H0810/16).

Although direct identifiers such as name and date of birth are not contained within the data, the amount of information about individual patients is quite detailed so it is treated as sensitive. Prior to data release, individual studies using CALIBER must be approved by the CPRD Independent Scientific Advisory Committee (ISAC). This study was granted individual ethical approval (protocol 14202_R).

CALIBER: Data collected

Primary care data: Clinical Practice Research Datalink (CPRD)

CPRD is an ongoing primary care database of anonymised medical records from general practitioners, with coverage of over 11.3 million patients from 674 practices in the UK (England, Wales, Scotland and Northern Ireland)(116). It represents one of the largest databases of longitudinal medical records from primary care in the world.

The population of active patients (alive and currently registered) on July 2013 was 4.4 million (6.9% of the UK population) and is broadly representative in terms of age, sex and ethnicity of the total UK population. The CPRD is therefore a rich source of health data for research, including data on demographics, symptoms, tests, diagnoses, therapies, health-related behaviours and referrals to secondary care.

Clinical encounters are entered onto the CPRD database using READ codes. READ terminology is a structured hierarchy of both medical and non-medical terms covering several areas including categories for signs and symptoms, diagnoses, investigations, treatment and therapies, drugs and appliances, occupations and administrative processes. They therefore offer a comprehensive list of clinical terms that can be used to describe the care and treatment of patients.

Information in the CPRD is recorded in several tables, which can be linked by the pseudonymised patient identifier to build up a complete picture of a patient's healthcare experience.

Patients – one row per patient, with demographic details such as year of birth, date of death and registration dates.

Practices – one row per practice, giving details such as region of the UK and the date when the practice achieved 'up-to-standard data' (Described further below).

Consultations – each patient episode is considered a 'consultation' and all data are entered in consultations (face-to-face, telephone or administrative). This table allows diagnoses and prescriptions entered in the same consultation to be identified.

Staff – one row per staff member, with gender and role.

Events –Each event is linked to a single consultation and an event date, a medical dictionary code (READ code), product dictionary code (Multilex) and/or associated information in free text.

Clinical – READ coded diagnoses entered by the GP with additional data such as observations

Referrals – referrals to secondary care, with the indication recorded as a READ code

Immunisations – records of immunisations

Therapy – prescriptions

Test – results of laboratory tests, each with a READ code

Secondary care data: Hospital Episode Statistics (HES)

Hospital Episode Statistics (HES) is a database warehouse containing details of all admissions, procedures, outpatient appointments and Accident & Emergency (A&E) attendances at NHS hospitals in England(109,111,113). It is a records-based system that covers all NHS trusts in England, including acute hospitals, primary care trusts and mental health trusts with information on each hospitalisation stored as a large collection of separate records (one for each period of care) in a secure data warehouse. This data is collected during a patient's time at hospital and utilised to allow hospitals to be paid for the care they deliver. Currently, in the CALIBER dataset, only information on admissions and procedures from HES is available.

Data on diagnoses are logged using the ICD-10 coding system whilst information on procedures is stored using the OPCS4 coding system(117,118). ICD-10 is the 10th revision of the World Health Organization's medical classification system(119). It contains codes for diseases, signs, symptoms, abnormal findings, complaints, social circumstances, and external causes of injury or disease. OPCS-4 is the coding system for operations, procedures and interventions performed during inpatient stays, day case surgery and some outpatient treatments in NHS hospitals (119). Patients are identified by their NHS number in the same way they are for CPRD.

Death registry & deprivation: Office of National Statistics (ONS)

The death registry for England and Wales curated by the Office for National Statistics (ONS) includes the date of death and the causes entered on the death certificate. A single underlying cause of death is allocated according to the WHO ICD-10 algorithm based on the information recorded on the death certificate, likely causal sequence and ICD selection rules (120).

Deaths in England and Wales have been coded using ICD-10 since 2001 and ICD-9 in previous years. Due to a change in the rules for selecting the underlying cause from ICD-9 to ICD-10, the causes of deaths are not directly comparable between 2001 onwards and previous years.

The Index of Multiple Deprivation (IMD) is a composite measure of deprivation calculated by ONS using indicators for super output areas (postcode areas). It covers the following domains; (i) Income, (ii) Employment, (iii) Health and disability, (iv) Education, skills and training, (v) Barriers to housing and services, (vi) Crime and (vii) Living environment (121).

Data management

See Appendix C CALIBER Data management - Figure C 1 Merging cohort data with co-variate data, Figure C 2 Creating main cohort/denominator file, Figure C 3 Creating platform files for analyses.

Since my objective in this study was to investigate regional variations in tonsillectomy rates in relation regional variations in the underlying disease burden for tonsillectomy, the ideal measure would have been a primary care measure of bacterial tonsillitis. However, there remains considerable heterogeneity between clinical signs and symptoms of all acute sore throat infections with no objective measure of bacterial tonsillitis currently available. There is disagreement even amongst GPs following history and physical exam on the diagnosis and coding of tonsillitis (105). Therefore, since I had no way to accurately and reliably measure tonsillitis in the primary care I chose to use a sensitive measure that included all sore throat consultations, assuming that the ratio of sore throat consultations to tonsillitis consultations did not vary from region to region.

Data were provided as comma separated value files, on a secure server. Files were separated by cohorts of pre-defined medical co-morbidities, two cohort files that had basic patient cohort information (GP practice ID, region, GP registration date, date of birth, gender, IMD

rank, ethnicity, date of leaving the practice, date of death), a file with all sore throat consultations (using codes defined in the process above), and a file with all tonsillectomies.

I undertook several steps that allowed these various files to be converted into the final six platform files that were ready for analysis. The steps are outlined in the flow diagram below and described as text in the following paragraphs.

Within the CALIBER cohort, that had patients from 1988 to 2010, there were 4,703,547 patients, 2,891,511 sore throat consultations and 29,578 tonsillectomies (not necessarily for recurring tonsillitis on each occasion) recorded. However, I restricted our cohort so that inclusion began from 1st January 1997, as this was the date HES data was linked to CPRD. Furthermore, I restricted our cohort so that each patient could only enter following the date their GP practice was deemed to be up to standard practice (UTS). This was to ensure high quality data were used for analyses. Approximately 400,000 patients were dropped as they did not meet these two basic requirements.

Sore throat consultations that occurred on the same day (104,362) or within 21 days of the previous sore throat consultation (210,423) were dropped to ensure I was measuring different illnesses. Had I changed the definition to 14 days or 7 days I would have removed 169,060 or 104,630 consultations, respectively. However, given that an average sore throat lasts for seven days (FluWatch) and a follow up consultation can be scheduled for two weeks after the first to ensure the symptoms have resolved, I felt that 21-day washout after the first consultation would ensure that review consultations could not falsely inflate the rate of consultations for active episodes of sore throat.

Whilst there were 29,578 tonsillectomies reported within CALIBER more than 1500 were duplicated entries due to the use of additional codes. Therefore, only 28,046 patients had undergone a tonsillectomy within our cohort. The sore throat consultation and tonsillectomy files were merged and then added to the cohort file that had dates of entry and exit from our cohort. Consultations and tonsillectomies that occurred on dates that were outside the period of risk defined in our cohort were excluded. Therefore, more than 1.14 million sore throat consultations and 10,764 tonsillectomies were removed. There were 3,560,864 patients in my study.

I added a variable that defined recurrence as having three sore throat consultations within a year (each sore throat consultation being 21days apart from each other) and restricted to tonsillectomies that had received an ICD-10 code for recurring tonsillitis. 16,618 had received a tonsillectomy for ICD-10 code of recurring tonsillitis.

Three files were created to separate the outcomes of interest (sore throat consultation, recurring sore throat consultations and tonsillectomy for recurring sore throat).

Files were further divided by age, into children (15 years and younger) and adults (16 years to 44 years old. This age range was chosen because above the age of 44 the risk of tonsillectomy is very low, however the number of years added to the denominator is substantial, resulting in a falsely inflated denominator. These six files formed the platform files from which our data analysis was undertaken.

CALIBER: data quality

Representativeness

When compared with the UK census in 2011 (122,123), CPRD patients are broadly representative of the UK population in terms of age and sex (109). Patients are also comparable to the UK census in terms of ethnicity(124), and comparable to the Health Survey for England for body mass index distribution in most patient subgroups(125). However, the CPRD may not be representative of all practices in the UK based on geography and size(126). There are also certain patient groups that are missing from primary care records, such as prisoners, private patients, some residential homes and the homeless.

Validity and misclassification

Validity in this context describes how well a specific code (or combination of codes) used pragmatically in routine healthcare records describes the presence of a specific diagnosis needed for research purposes. Conversely, it also relates to how well the absence of a code (or a combination of codes) in the routine healthcare record predicts the absence of a specific diagnosis needed for research purposes. CALIBER researchers need to rely on READ codes that GPs have assigned to consultations or ICD-10 codes that hospital coders have assigned to hospitalisations. In routine practice, clinicians may not have applied the strict case definitions when allocating diagnostic codes that are defined by researchers. If this happens it is termed misclassification.

Misclassification can occur in the selection of the patient population, exposure, confounders and outcomes. It is commonly categorized as either non-differential or differential misclassification. For example, in non-differential exposure misclassification, the misclassification is deemed unrelated to the occurrence or presence of disease. In contrast, if the misclassification of exposure is different for those with and without disease, it is differential. Misclassification can have a large impact on how well we can predict for a disease from the code, and therefore affect the sensitivity, specificity, positive and negative predictive values. Even small changes in sensitivity and specificity can have large impact on the calculation of outcome incidence(127).

Studies that have explored the validity of diagnoses on CPRD have shown high positive predictive values. In fact, a systematic review of CPRD validation studies noted that diagnoses were generally reliable(128). And, where evaluated these studies have also shown an incidence of disease that is comparable to other UK datasets(129-133). However, these studies rarely describe other components of validity such as negative predictive value, sensitivity or specificity.

Numerous epidemiological studies have been performed using HES with validation studies confirming that HES records on RTIs appear to be both reliable and complete (108). However, due to its primary purpose as an administrative dataset, caution is advised when undertaking studies due to its research utility being limited by data accuracy at illness level and limitations of the ICD-10 coding system. In fact, an audit study investigating the validity of codes in HES for surgical patients, when compared to patients' complete hospital records, demonstrated that at least one change was required in 55% of cases. However, only 17% of primary diagnoses required changing(134). When looking more specifically at HES coding of surgical procedures, 12% of coded operations required revision(135). In a similar study where the authors looked more specifically at HES coding for otolaryngological admissions and procedures they reported that primary diagnoses and procedures were incorrect 13% of the time (136). These studies were primarily aimed at reporting lost hospital earning through inaccurate coding and may have a different priority when suggesting primary codes for diagnosis and operations. Additionally, personal correspondence from the authors of these studies has shown that coding error increased with complexity of admission and operations. Indeed, there was little coding error for patients admitted electively for a tonsillectomy (personal correspondence). Therefore, the evidence suggests good validity and little misclassification in the use of tonsillectomy codes to depict patients who underwent tonsillectomy.

The implications of CPRD misclassification in our study can be seen in the following two examples: 1. Since the use of antibiotics in primary care is now highly scrutinised due to bacterial resistance, it is possible that General Practitioners wishing to prescribe antibiotics for their patients with upper respiratory infections may be more likely to code their consultation as bacterial tonsillitis. This misclassification of severe viral upper respiratory

infection may reduce the specificity in detecting patients with bacterial tonsillitis. 2. Alternatively, GPs may code all patients with upper respiratory tract infections (including tonsillitis) as 'sore throat', for ease. Using only the bacterial tonsillitis code to define tonsillitis would miss cases and reduce sensitivity. Indeed, there is evidence that there is considerable variation amongst coding practices for GPs, especially for sore throat illnesses (137,138). To reduce the risk of misclassification it was important to have an appropriate definition of my outcome of interest (phenotype) and a robust method that would allow the identification of READ codes that identified my outcome of interest (phenotyping). This process is detailed below.

Sore throat consultation phenotype

Since my objective in this study was to investigate regional variations in tonsillectomy rates, and the underlying disease burden for tonsillectomy, variations in coding practice could greatly influence my conclusions. The underlying disease burden in this case related to recurring tonsillitis (57). However, we know this that there is variation in how GPs diagnose tonsillitis (105) and how they define recurrence of tonsillitis (139). Therefore, if a very specific definition of underlying disease burden was used (e.g. "recurrent streptococcal tonsillitis") then variations in coding practice may be more readily interpreted as variation in underlying disease burden. However, if a very sensitive definition of underlying disease burden was used (all codes that could be related to recurrent tonsillitis, such as 'sore throat') then variations in coding practice would be diluted, but conclusions may be harder to reach. Therefore, I chose to use sore throat (which includes tonsillitis) variation as a surrogate marker for tonsillitis variation, which assumes that the 'ratio' of sore throat to tonsillitis is the same across the geographical entities you compare. It is an assumption which is however sensitive to help-seeking for sore throat being the same across the land. Additionally since here is higher number of sore throat consultations than tonsillitis consultation modelling is relatively more robust to small number issues than it would have been if I had gone out for tonsillitis.

I opted, in the end, to use a sensitive measure of disease burden underlying tonsillectomy, which describes the overall rate of all consultations for sore throat infections, rather than being specific to bacterial tonsillitis, as it made fewer assumptions on the data. This is not

unreasonable, as the ratio of sore throat to tonsillitis is likely to be the same across the various geographical regions which are compared in the thesis. The sore throat phenotype also includes tonsillitis. This meant that the specificity of my case definition – **acute sore throat consultations** - was low when measuring bacterial tonsillitis resulting in many non-true cases being included, potentially diluting variations in the recurrent tonsillitis. Having an increased number of sore throat consultations included may be beneficial, however, as it results in my modelling being more robust to small number issues and variation due to chance. To investigate if my definition of sore throat infection was valid I cross referenced the incidence of sore throat calculated using this phenotype definition CALIBER and compared it the predicted incidence of sore throat to my population study that had detailed information about rate of consultations for sore throat illnesses.

Recurring sore throat phenotype

Whilst our definition of sore throat consultations was broad and not specific to consultations of bacterial tonsillitis, I decided to opt for a more specific definition of recurring sore throat, which may relate more accurately to patients who were at risk of a tonsillectomy. National guidelines advocate referral to ENT to consider a tonsillectomy if a patient has 3 episodes of tonsillitis a year for 3 years, 5 episodes a year for 2 years or 7 episodes a year for one year(57). However, I was cognisant that not every sore throat requires primary care consultation, especially now that national guidelines discourage the overuse of antibiotics for throat infections(140). Therefore, GPs and ENT surgeons often judge the frequency of sore throat recurrence based on the patients' recall rather than documented evidence of consultations. I undertook a sensitivity analysis to assess the impact of defining recurring sore throat as 3 or more versus 5 or more sore throat consultations over a year. I assessed which definition could sensitively and specifically detect those that had a tonsillectomy for recurring tonsillitis. Whilst both definitions had high specificity, defining recurrence at 5 or more sore throat consultations in one year produced an extremely low sensitivity (3.5%) (see Table 5 Sensitivity analysis for definition of recurring sore throat). I therefore, chose to define recurrence at 3 or more sore throat consultations within a year, with each consultation being more than 21 days apart from each other. Our population study of sore throat showed the median sore throat

illness lasted 7 days (interquartile range 4-10days, and therefore 21 days was considered more than sufficient to define a second illness.

	Tonsillectomies Undertaken in this group	Tonsillectomies undertaken not in this group	Total patients who meet definition of recurrence	Sensitivity	Specificity
>2 episodes a years	7040	9578	45,443	42%	98.9%
>4 episodes a year	585	16033	1260	3.5%	99.9%

Table 5 Sensitivity analysis for definition of recurring sore throat.

This table shows that using a definition of more than 2 GP sore throat consultations (based on READ codes on CPRD) more sensitively identifies patients who had a tonsillectomy for recurring tonsillitis (based on ICD-10 codes on HES) than 3 or more GP sore throat consultations (n=45,443).

Tonsillectomy phenotype

The majority, but not all, tonsillectomies are undertaken for recurring tonsillitis. The procedure is also indicated for cancer and obstructive sleep apnoea. Therefore, it was important to identify patients who had tonsillectomy for recurring tonsillitis. Additionally, all tonsillectomies in England are undertaken under anaesthesia and in a secondary care setting. Patients who have a procedure receive an OPCS4 code and the indication for their procedure is recorded as their diagnoses with ICD-10 codes. For these reasons, I used the HES database to identify patients who had undergone tonsillectomy, and then restricted patients to those who had received an ICD-10 code for tonsillitis as their primary diagnosis.

Phenotyping code lists

As part of the phenotyping process, codes of interest (*e.g.,* diagnostic, symptom, medications) needed to be identified and listed according to the relevant source dataset terminology (*e.g.,* selecting all READ terms for sore throat infections to assess sore throat infection consultations in primary care). There are thousands of potential terms per terminology, with a varying number of terms (handful to hundreds) required depending on what the disease state is.

Code lists were developed by an iterative process which involves searching for terms in the dictionaries, combining selections of terms to derive a final set of chosen terms, and assigning a category to these terms.

In CALIBER, the production of code lists is assisted by the use of; (1) the CALIBER Data Portal, a web portal for researchers to access descriptions of contributed CALIBER clinical phenotypes (phenotypes), the underlying development process and code lists of Read, ICD-10 or OPCS codes used to define them and (2) R packages created by Dr Anoop Shah; the R CALIBERcodelist package(141). This is illustrated in the Figure 7 Process for generating a code list using the R CALIBERcodelists package.

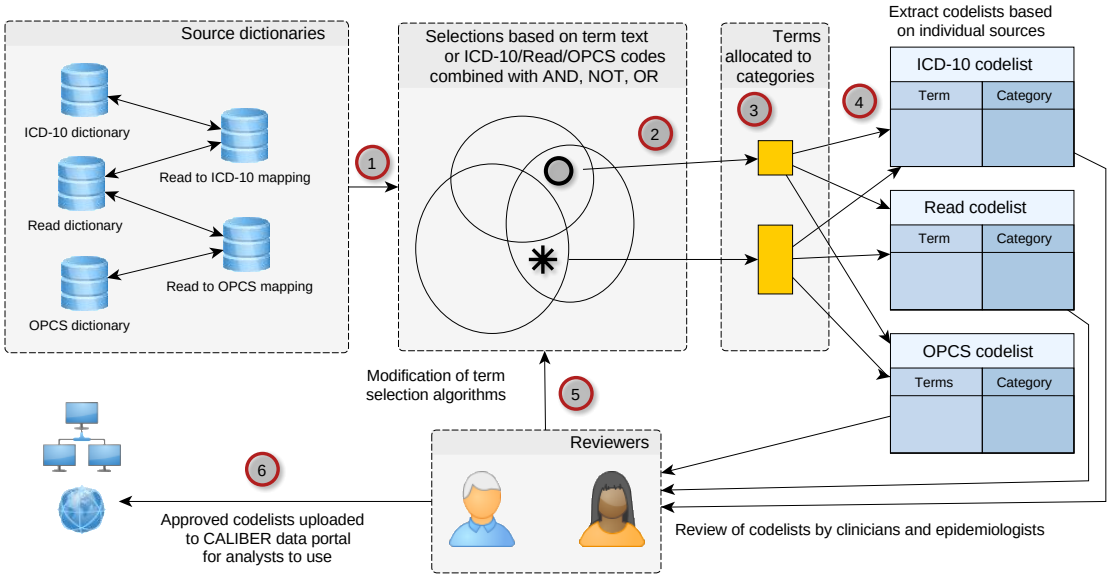


Figure 7 Process for generating a code list using the R CALIBERcodelists package

This diagram demonstrates how multiple codes, across code libraries, received during routine medical care can be phenotyped to reliability represent a medical condition.

1. Decide which source dictionaries to use, e.g. READ, ICD-10 and/or OPCS
2. Create a selection of terms from source dictionaries (*e.g.,* all terms containing the word 'throat' or all ICD-10 codes beginning with 'I10'). Combine selections using Boolean operators such as AND, OR or NOT to identify exactly which terms are of interest.

3. Allocate a category to terms in a particular selection, *e.g.* 'Sore throat'
4. Set the metadata for the code lists under construction (version number, category descriptions, author name, date). Extract the READ, ICD-10 and OPCS code lists and save in a standard format.
5. Undertake a consensus meeting with clinicians and epidemiologists to review the code lists and suggest any changes. The algorithm may be iteratively changed with results compared with previous versions if necessary.
6. Approved code lists with documentation can be shared on the CALIBER data portal for use in subsequent studies

In CALIBER, sore throats can be classified using either diagnostic or symptom codes, each of which has potential limitations:

Diagnoses codes: Assuming the patient and/or healthcare professional know what type of sore throat the patient has, it can theoretically be coded using the specific READ or ICD-10 code. However, some codes are ambiguous, for example 'throat pain' could have infectious, inflammatory, neuropathic or traumatic causes. In our case, I would only be concerned with infectious causes of sore throat, which if they were to become recurrent a tonsillectomy may be recommended. In addition, some patients may have conflicting diagnostic codes issued on the same day.

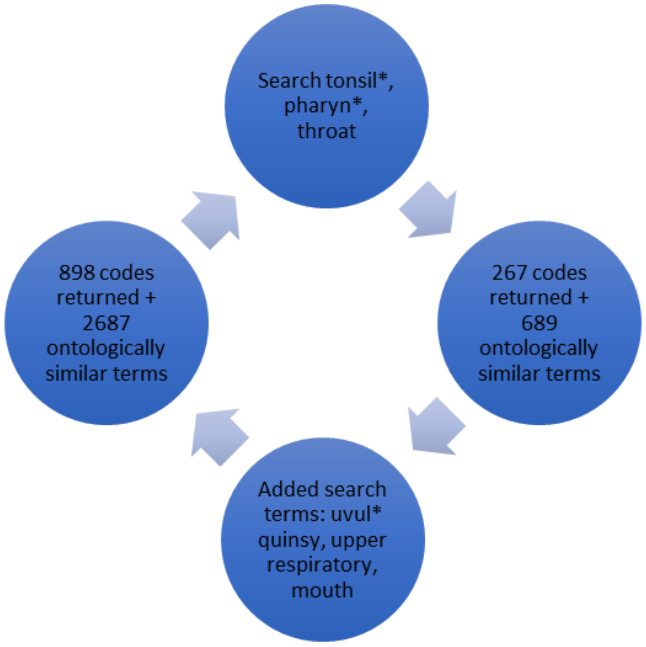
Symptom codes: In contrast to diagnostic codes, these are usually more ambiguous and may range from 'sore throat' to 'cough' and 'fever'. Based on clinical knowledge, sore throat in a child can be classified as either laryngitis or tonsillitis (*i.e.,* URTIs). In contrast, cough could either be an URTI or LRTI whilst fever, a sensitive marker for an infection, is more non-specific and may be a possible RTI.

In developing the algorithm, I reviewed a previously published code list for identifying sore throat infections in UK primary and secondary care datasets(2,142,143). I used this list to add to my search terms and as final check against my code list to ensure no codes had been omitted.

Diagnostic and symptom codes were searched through the READ and ICD-10 dictionaries using the R CALIBERcode package. The primary search terms "throat" "tonsil*" "pharynx*"

"larynx*" were used. The programme returned all codes that contain the search term as a prefix or suffix as well as codes that were ontologically similar.

The initial search with these terms returned 267 codes that contained the search terms and a further 689 that were similar ontologically. Review of the ontologically similar codes provided additional search terms: "uvul*", "quinsy", "fauces" "upper respiratory" and "mouth". The search was reinitiated with both the primary and additional search terms. This returned 898 codes that contained the search terms and 2687 ontologically similar. The ontologically similar codes were initially searched to ensure there were no missed terms that could be used on a further search. Since there were no new search terms the iteration was stopped and the full code list was given to two reviewers.



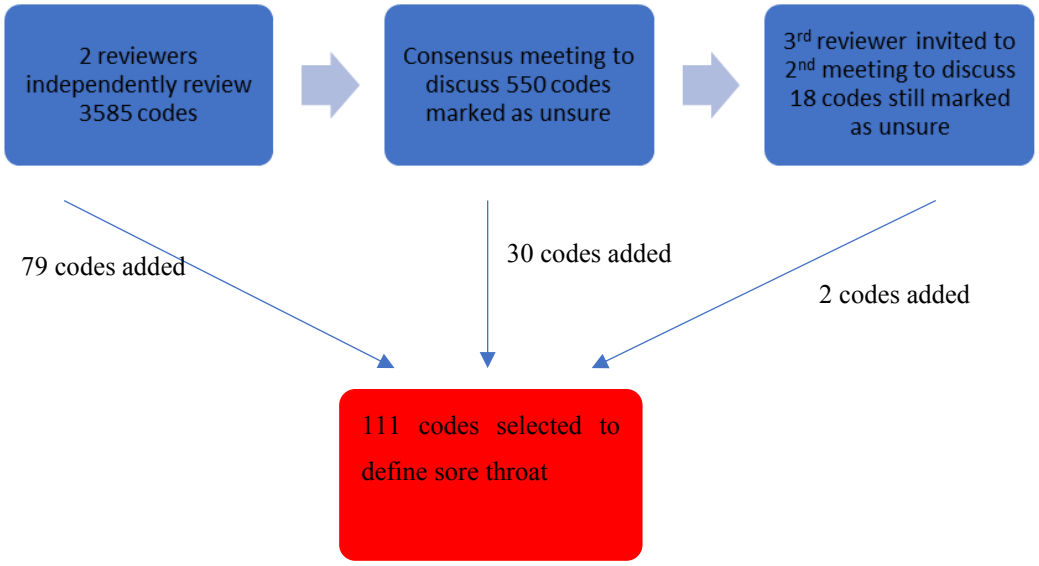


Figure 8 Generating a sore throat codelist

. This diagram shows I searched for codes of interest using key search terms that were revised and search reiterated based on results. It also shows how from 3585 codes I was able to identify 111 codes to phenotype for sore throat infection.

The reviewers (NM, LM and MM) were asked to short list the extensive list to codes that could be used as a marker for consultations that related to acute sore throat infections, that could be treated with a tonsillectomy should they become recurrent. They were allowed to mark codes as *definite exclude*, *definite include* and *unsure*. The task was undertaken on excel and once complete the two files were merged. All codes that both reviewers had marked as definite exclude were dropped (2956 codes excluded). 79 codes were marked by both reviewers as definite include. All codes where reviewers disagreed, or were marked by either reviewer as unsure disagreed were discussed. 550 were marked as unsure by one or both reviewers, or were marked differently by reviewers. Consensus was reached between reviewers on 532 codes, with a 30 further codes added to the list of inclusions. For eighteen codes where consensus could not be reached a third senior ENT doctor was asked for his opinion, and codes were included or excluded depending on the majority vote (2 or more reviewers). The process is illustrated in Figure 8 Generating a sore throat codelist. For a full codelist please see Appendix U - Sore throat codes.

These code lists were subsequently discussed and finalised at a consensus meeting that consisted of an ENT surgeon (Anne Schilder), a biostatistician (Hannah Evans) and an infectious diseases epidemiologist (Andrew Hayward). One hundred and eleven codes were classified as either "Probable" (n=30) or "Possible" (n=81), in relation to whether they described codes for consultations about acute sore throat infections, that could be treated with a tonsillectomy should they become recurrent. Codes that were "Probable" accounted for less than 10% of all codes used and had very low sensitivity (<4%) for detecting patients who went on to receive a OPCS4 code for tonsillectomy associated with an ICD-10 code for recurring tonsillitis. Additionally, it was known that there was heterogeneity in the coding of sore throat illnesses by GPs (105). And finally, tonsillectomy may actually be of benefit for recurring sore throat infections, other than bacterial tonsillitis. For all of these reasons, I chose to use a sensitive code list of sore throat consultations – that is all 111 codes.

Analysis of the FluWatch dataset showed that median sore throat illness lasted 7 days (interquartile range 4-10days), and therefore 21 days was considered more than sufficient to define a second illness. This ensured that I did not mislabel surveillance checks on previous illnesses as a new throat infection. I undertook a sensitivity analysis using a 7, 14, and 28 day wash out period, but since there was no change in the overall incidence I chose the more conservative period of 21 days.

This procedure was repeated for codes that would relate to tonsillectomy on OPCS4 coding and tonsillitis related diagnoses on ICD-10. This procedure was relatively simple as there were only 3 OPCS codes for tonsillectomy (F34..) and 8 (J03/J35) ICD-10 codes for tonsillitis related conditions.

Categorising co-variates

Age was categorised into preschool (0-4 years), early school (5-15) adolescence and young adult (16-24), early adulthood (25-44). Adults over the age of 45 were rarely at risk of recurring sore throat (only 2% of those with one sore throat consultation go on to have 2 more within a year) or tonsillectomy for recurring sore throat (less than 10% of all tonsillectomies, but more than 50% of the time at risk), therefore, I removed them from analyses as I felt that would artificially and inappropriately inflate the denominator.

Age category (years)	Total person-years at risk	Number of sore throat consultations	Patients with recurring sore throat *n* (% of those who had sore throat)	Total number of patients who were coded as having tonsillectomy for recurring tonsillitis	Mean (95%CI) number of sore throats in those coded as having tonsillectomy for recurring tonsillitis	% of those coded as having a tonsillectomy for recurring tons who had less than 3 GP sore throat consultations
0-4	347,734	117443	4871 (6%)	1254 (36%)	6.09(5.89-6.28)	11
5-15	2,253,956	358790	13385 (9%)	6396 (45%)	4.67(4.58-4.76)	35
16-24	2,360,050	267991	12,726 (7%)	4274 (38%)	4.30(4.20-4.39)	35
25-44	5,347,471	432555	10,381 (4%)	3490 (29%)	3.57(3.48-3.66)	41
>44	$1.06_{10}7$	263223	4080(2%)	1204 (23%)	2.27(2.17-2.37)	68
Total number	$2.10_{10}7$	1,440,002	45,443	16,618	4.28(4.22-4.33)	37

Table 6 The role of age in recurring sore throat and tonsillectomy.

This diagram demonstrates the number of sore throat consultations as recorded on CPRD compared to tonsillectomies for recurring tonsillitis recorded on HES by age group. As can be seen those over 44 account for half of the time at risk but less than 10 percent of all tonsillectomies, (n=1,440,002 person years).

Ethnicity was defined by ONS census categories into White British, Indian, Black African, Black Caribbean, Black other, Bangladeshi, Pakistani, Other Asian, Chinese, Mixed, Other, Unknown. Participants' geographical region was categorised by the 10 geographically based Strategic Health Authorities that manage local health care: North East, North West, Yorkshire & The Humber, East Midlands, West Midlands, East of England, South West, South Central, London and South East Coast. Social deprivation (categorised into national quintiles based on the index of multiple deprivation: IMD1 describing the most deprived quintile and IMD5 describing the least deprived quintile).

Code lists for the following conditions (that is potential risk factors for sore throats, recurring sore throats and tonsillectomy) were developed by other researchers at the Farr Institute of

Health Informatics using the same methodology I have used to develop the sore throat code list. Code lists were stored on a secure server with a file that explained how they were developed and the final do file that could be run to generate them again. Below is a list of how I categorised other potential risk factors for analysis.

Respiratory illness was classified based on published literature as a child having ever been given the diagnosis of asthma (any code starting H33) (144) or an adult having been given the diagnosis of asthma (any code starting H33) (145) or chronic obstructive pulmonary disease (a specific COPD code, more than one prescription of a COPD medication and spirometry)(146). Obesity was classified based on published literature (125) and was defined on body mass index greater than 29, recorded at least once during their window of observation. Obstructive sleep apnoea was diagnosed for any patient that received codes for this condition during their window of observation. HIV status was classified as positive if the patient had ever received the code. I did not use AIDs defining illness to define HIV status in the absence of a HIV positive test result. Eating disorder was defined based on published literature(147). Cases were identified if they had received the diagnosis of "Anorexia Nervosa", "Bulimia Nervosa" or "Eating disorder unspecified", "Atypical AN" or "Atypical BN" during their window of observation. Alcohol consumption was measured in adults only and based on previously published literature(148): Non-drinkers were patients who had codes such as "teetotaller", "non-drinker", "stopped drinking alcohol" and "ex-drinker", Mild-moderate drinker was used to classify those that had received codes such as "drinks rarely", "drinks occasionally", "alcohol intake within recommended sensible limits", "light drinker" "drank within daily and/or weekly recommended sensible drinking limits for the UK", and heavy drinkers were classified as those who received codes such as "alcohol intake above recommended sensible drinking limits", "hazardous alcohol use" and "exceeded daily and/or weekly sensible drinking limits." **Smoking status** was defined based on published literature into non-, ex- and current smoker(149). **Diabetes**: Patients were classified as being diabetic if they received at least 1 prescription for a noninsulin antidiabetic drug (NIAD) (150) during their window of observation. **Hypertension:** Patients were classified as being hypertensive if they were diagnosed by their GP as being hypertensive, had 3 more high systolic or diastolic

blood pressure measurements within one year, or patient received 2 blood pressure lowering medications within one year (113).

Completeness and missing data

Whereas misclassification is important for the case definition, missing data are more relevant to co-variates in the study. Missing data are common in large datasets, collected over long periods of time (151). Additionally, when a dataset is created for record keeping in busy clinical practice, missing data may be more likely than in cohort studies that are dedicated to the creation of a research-ready database (152). Since CPRD is the largest database of primary care in the world, collected for routine medical records, missing data are expected. There are two mechanisms that keep missing data down. The first is routine data quality checks: Practices participating in CPRD are assigned an "up-to-standard" date by CPRD custodians based on acceptable standards on ten practice-based measures of quality, completeness and representativeness. Once deemed "up-to-standard", their data is marked as suitable for longitudinal data research. Conventionally, clinical data from patients are restricted from the date their practice were deemed "up-to-standard". Practice data are checked once delivered to CPRD for data quality issues. Any practices submitting poor data is provided feedback and if coding practices are not rectified, data from their practice are subsequently removed from CPRD (114). The second relates to the introduction in 2004 of a national incentive payment program for GPs that encouraged better record keeping of key data items (for example smoking status and the delivery of services to key patient groups). This resulted in a large drop in missing data (109).

Nevertheless, the issue of missing data needs careful consideration as incorrect assumptions of their underlying mechanisms could have a big impact on the conclusions reached. Data can be classified as missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR) (153). Three examples are shown below of how the different mechanisms of missingness could occur in a study examining tonsillitis rates through patients who received READ codes of tonsillitis in CPRD: 1. Tonsillitis codes could be MCAR if all CPRD files were lost by accident for a random day. Patients who are missed from the cohort by

complete chance, and not based on their own characteristics or the way tonsillitis is coded. 2. Tonsillitis codes could be MAR if they were recorded more frequently in children than in adults, as adults frequently present with more than one symptom. Therefore, in adults, tonsillitis cases were missed at random. In this case, the chance of missing cases of tonsillitis did not depend itself on the presence of tonsillitis but on another patient characteristic – i.e. age. 3. If adults with tonsillitis were more likely to attend Accident and Emergency, or walk-in centres, then cases missed from CPRD could be described as MNAR. Unfortunately, the assumptions that differentiate the mechanisms of missingness are difficult to ascertain. Multiple strategies exist for reducing bias produced by missing data and be divided into deletion methods and imputation methods.

Deletion methods

There are two deletion methods available, listwise and pairwise deletions. Both are more applicable if the data are assumed to be missing completely at random.

Listwise deletion or complete case analysis restricts analyses to patients who have data available on all variables, i.e. have no missing data. This allows for comparability across analyses, but reduces statistical power and produces biased results if data are not MCAR, e.g. investigations into the relationship of age and tonsillitis would be restricted to patients who have no missing data on any variables (including smoking). If, however, smoking status was more likely to be missing in adults than in children and all patients with missing smoking status were excluded, the association between age and tonsillitis would be biased.

Pairwise deletion or available case analysis restricts analyses to cases in which variables of interest are present. This allows maximum use of data available, but prevents comparison across analyses as each analysis related to a slightly different sample, e.g. investigations into the relationship of age and tonsillitis, would be restricted to cases that had no missing data for tonsillitis or age only. Patients who had missing smoking status but had data on age and tonsillitis would still be included. Whilst this method makes fewer assumptions on the mechanisms of missingness and allows the use of as many cases as possible, we cannot compare across analyses as a different sample was used in each analysis.

Imputation methods

Imputing, or replacing, missing data can be divided into single or multiple imputation methods. Imputing methods are especially valuable when data are assumed to be missing at random.

Single imputation methods include replacing missing data with the nearest value(154) or field average (mean, mode, median)(155). This method can be supplemented with adjusting for missingness, by creating a variable that is positive when the data field was initially missing. These methods reduce variability within the variable and may weaken correlation between the imputed variable and others. Alternatively, missing field can be replaced by a regression value, or one that is predicted based on the relationship of the outcome of interest and a second variable. For example, if height varied with age and one height recording was missed in a child's clinical records, the missing field could be imputed based on a regression model between the observed correlation between his height and age. However, this method overestimates model fit and weakens variance. Overall, single imputation models introduce false precision, by reducing the obvious uncertainty of the missed variable, and they ignore associations of the missing variable with the whole dataset, therefore, they are not ideal for inference.

Multiple imputation methods, builds on regression imputation, and involves filling in each missing value with several plausible values, based on many other variables, in a way that reflects the uncertainty about them and their relationship to the overall database. This results in the creation of multiple datasets, which are each analysed individually, and the estimates of the missing values are combined to produce a likely value for the missing field. This method allows for uncertainty inherent to the variables' missingness and a more accurate measurement of the relationship of the missing variable to other variables in the data set. However, multiple imputation greatly adds to the complexity of the modelling and there is considerable room for error in the manner the model is specified(156).

In our analysis co-variates appeared to be missing at random after adjusting sore throat incidence for age and sex. Previous studies using CALIBER have assumed co-variates are MAR

following basic tests (Associations between polymyalgia rheumatica and giant cell arteritis and 12 cardiovascular diseases(157,158)).

I used a mixed approach to deal with missing covariates since most variables in the full dataset had a degree of missingness. I felt that within our cohort of children and young adults (0-45 years old) a missing code was more likely to represent the absence of a disease rather than missed diagnosis for conditions that were not actively being routinely measured in our population (e.g. atrial fibrillation, heart failure, hypertension, multiple sclerosis, cirrhosis, cardiac valve abnormalities, endocarditis, Parkinson's, COPD, diabetes, dialysis, liver pathology, renal disease, respiratory disease and heart valve diseases). It would not be appropriate to impute diagnoses of these chronic conditions as I did not believe they were missing at random (as absence of disease is not actively recorded).

Obesity was considered a key co-variate. Obesity was strongly associated with obstructive sleep apnoea, which can be another indication for tonsillectomy, mostly in children but also occasionally in adults as part of more complicated airway surgery. Additionally, obesity can also be associated with gastro-oesophageal reflux disease that results in non-infectious sore throat. For these reasons it was considered an important condition to control for in the assessing the impact of sore throat infections on tonsillectomy. Smoking was considered as key co-variate as smoking can be related to poor oral hygiene and increased risk of infections. though to be Missing data in other co-variates (obesity and smoking status – see Table 7 Table of missingness in key co-variates.) were evaluated through a sensitivity analysis comparing the effect of complete cases and best case (missing is absence of disease) and imputation on overall model. Multiple imputation was implemented using the M algorithm in the statistical package Stata 13.1, to replace missing values in exposure and risk factor variables. Ten multiply imputed datasets were generated, and Poisson models were fitted to each dataset. Coefficients were combined using Rubin's rules. The Kolmogorov Smirnov test was used to compare the distribution of observed versus imputed log transformed covariates.

Results from analysis evaluating risk factors for sore throat from multiply imputed dataset were compared to those created from complete cases and best case. Since the direction of the effect of the covariates on the outcome was unaltered between analyses I chose to analyse obesity as best cases (i.e. absence of code implied absence of disease). Whilst GPs

were unlikely to record BMIs for all of their young patients I assumed that they would if their patient was obese.

Since the effect of smoking on sore throat was altered depending on whether I used complete case or best case I decided to use complete cases as I felt it did not assume that patients with missing codes were non-smokers, and the direction of the effect was similar to the imputed dataset.

	% Missing
Obesity	94
Smoking status	31
IMD score	1

Table 7 Table of missingness in key co-variates.

This table shows the percentage of patients in CALIBER dataset that have missing codes for key co-variates: obesity, smoking status and IMD.

Summary

In summary, CPRD and thus CAILBER is representative of the UK national population, but is missing certain groups such as the homeless and prisoners. Validity in CPRD and HES is generally good and misclassification is low. For our study, I chose a sensitive definition to calculate underlying disease burden of tonsillectomy (i.e. all sore throat infections) to reduce the impact of misclassification and variation in coding by GPs. Completeness of data in CALIBER is better following the introduction of national data recording quality incentives. Missing data for co-variates are most likely missing at random and therefore, I undertook multiple imputation to investigate the impact of missingness in our analyses. The multiple imputation dataset was used to investigate co-variates associated with sore throat and tonsillectomy. Sensitivity analysis showed similar results between multiply imputed dataset and available case dataset. Therefore, the systematic component of variation was calculated on the available cases dataset to reduce the number of assumptions that are made.

CHAPTER THREE: Is there regional variation in the incidence of sore throat or help seeking behaviour?

Chapter synopsis

In this chapter I have presented the results of a longitudinal population based study of self-reported sore throat and help seeking-behaviour. This study determined the incidence of sore throat more robustly than ever before at 4.22 sore throat episodes (95% CI: 3.86-4.61) per 1000-person-days. Specific to children, population density, age and the presence of chronic illness were identified as predictors of sore throat in the community. Population density, age, duration of illness and fever were identified as predictors of help seeking behaviour for sore throat. For adults, age, gender, household size and smoking status were identified as predictors of sore throat, and EQ5D index score the only predictor of help seeking behaviour for sore throat. Importantly, this study demonstrated that whilst there is regional variation in sore throat incidence, both for adults and children, once the above factors are accounted for, regional variation becomes non-significant.

Introduction

Whilst previous epidemiological studies have shown considerable variation in tonsillectomy rates, both between countries (159)) and regions within a country (16,17,19,20), they have failed to investigate the incidence of sore throat, or help seeking behaviour (HSB) for sore throat (primary care consultation), across those same regions. In part, this is related to the difficulty in capturing acute illness information at population level on a large enough scale to inform our understanding of regional healthcare variations. Therefore, it is uncertain whether the variation in tonsillectomy rates are a warranted response to the regional disparities in disease burden (community sore throat infections) and HSB (GP consultation for sore throat), or whether there is systematic bias between regions that exposes patients to inequity in the management of recurring sore throats. The only previous UK based population study of sore throat was conducted on 198 pregnant women more than 3 decades ago, and it was conducted in Lambeth only. The study was not powered to assess predictors of sore throat. And whilst this study did assess help-seeking behaviour for sore throat (i.e. if the patient went to the GP), it was not on a large enough scale to accurately calculate predictors of GP consultation, and due to the local nature of the study could not assess regional differences in GP consultation. Therefore, there are no generalisable data about the incidence, predictors or regional differences in community sore throat. Nor are there data on the rate, predictors and regional differences in GP consultation for those with sore throat.

Objective

The objective of this chapter was to quantify regional tonsillectomy rate variation in relation to regional variation of self-reported sore throat and help-seeking behaviour in the community (objective A).

Research questions

I addressed this objective through the following 8 questions:

1. What is the incidence rate of sore throat in the community?
2. Is there a regional difference in the incidence of sore throat?
3. What are the predictors of sore throat in the community?

4. How much variation in sore throat exists once these population characteristics are accounted for?
5. In people who have a sore throat, who seeks medical advice?
6. Is there a difference in the rate of GP consultation in those with sore throat between regions?
7. What are the predictors of GP consultation in those with sore throat?
8. How much variation in GP consultation persists once disease and population characteristics are accounted for?

Methodology

The analyses used the FluWatch dataset described in Chapter 2. I have provided a detailed explanation of the FluWatch cohort, variables and missing data in chapter 2. Dependent and independent variables were defined in line with published evidence around sore throat. Missing data were managed by only analysing data for weeks where the household survey lead had responded with a completed weekly illness report.

Weighting

Since the survey was oversampled in the Southwest of England and under-sampled in those between 0-15 years I weighted analyses to age and regional structure of England to give locally and nationally representative estimates. The final weight also accounted for the method of sampling through households (that is participants from a larger household had a greater chance of being sampled compared to those from smaller households). Models used to estimate the incidence of sore throat and GP consultation rate for sore throat were weighted to make survey data more nationally representative.

Modelling

Analyses were undertaken separately for children (15 years and younger) and adults (16 years and over), as the epidemiology and risk factors for sore throat in these populations was known to be different.

Outcome Variables

On the first analysis I used self-reported sore throat as my main outcome variable. This outcome has been defined earlier in chapter 2. But in summary it relates to any self-reported moderate-severe sore throat over at least two consecutive days, with an associated fever in the absence of a cough.

Since participants could report more than one sore throat during their time at risk, logistic models were less appropriate. General linear models (GLM) were used to quantify the rate and predictors of sore throat, as they allowed repeated counts and could account for disparities in the time at risk. I assessed how different models predicted dispersion of the data and the participants that never reported sore throats. Whilst the mean and variance were similar, negative binomial models could predict variance more closely to the observed data than Poisson models. Poisson models predicted the rate of participants who never reported a sore throat better than negative binomial models. However, both Poisson and Negative Binomial multivariable models showed no difference in the direction of variable effects and therefore, Poisson models were used in the sections that follow. All analyses were conducted on the whole FluWatch cohort, children (n=1414) and adults (n=5946), separately.

On the second analysis I used sore throat GP consultation as my main outcome variable. This outcome was defined as self-reported GP consultation for sore throat infection. Since help seeking behaviour data were only available in the final three influenza seasons (May 2009-March 2011), analyses for rates and predictors of GP sore throat consultation were restricted to these periods. Logistic models were used to describe predictors of help seeking behaviour during self-reported sore throat illnesses (a binary outcome). All analyses were conducted on self –reported sore throat illnesses (defined above), for children (n=433) and adults (n=1760) separately.

As participants could report more than one sore throat or health care contact I clustered data to the level of the participant in all analyses. Conceptually, it seemed logical that patterns of disease may be clustered by region. To investigate this I also clustered data at the level of the GP. A multi-level model, clustering at patient and GP level, was found to more accurately represent that data compared to a clustering at patient level only (likelihood ratio test).

Therefore, I used multi-level (patient and general practice level) models to evaluate clinical and sociodemographic determinants of sore throat and GP consultation amongst those with sore throat.

Exposure Variables

Exposure variables are defined in chapter 2. In summary, I used age, ethnicity, participants' geographical region, population density (urban versus rural), quintiles of index of multiple, Influenza vaccination status, and health utility/status index and subdomains scores.Exposure variables that were associated with the outcome with a *p* value of <=0.1, on univariable analysis, were considered for inclusion in multivariable regression models with an *a priori* decision to include age and gender regardless of the association. Additional variables were included initially starting with variables that had the most number of plausible mechanisms of action and then adding those with fewer plausible mechanisms (as per our conceptual model described in chapter 2). Akaike's information criteria (AIC) (160) for sequential models was noted and the probability that each model could reduce the information loss as compared to the model with lowest AIC was calculated. Exposure variables were added sequentially and hierarchically, as random effects to the appropriate level, if they improved the model fit. The model was retained if it had a high probability of reducing information loss. I undertook tests for interaction between variables if both variables were independently related to the outcome and had a biologically plausible interdependent relationship to outcome. Once a full multi-level multivariable model had been created, region was added to the model to see if geography was still important after accounting for population and disease level predictors of sore throat, recurring sore throat and tonsillectomy rates.

Models developed to predict the incidence of sore throat and GP consultation rates were used to calculate the systematic component of variation (SCV). The SCV measures the degree of variation between regions. Other methods, such as extremal quotients, standard deviation and coefficient from variation were not used because they are greatly affected by differences in population size between regions. The SCV is an adaptation of the proportional hazards model. The method described(161) subtracts the random error (estimated through generalised linear models above) from the estimated total variance to calculate the systematic component of variation. This measure allows comparisons of variability between

regions but makes few assumptions about the nature of the variation and allows for appropriate amounts of sampling variation in the data. It has been suggested that variations giving SCVs greater than 3 are likely to be due largely to differences in practice style or medical discretion, and that high variation is described by a SCV of between 5.4 and 10.0, with SCVs greater than 10 being very high variation (162). Bevan et al (163)identified high variation as healthcare resource groups (HRGs) with an SCV greater than 6.6, the SCV for hip replacement.

All statistical analyses were undertaken on Stata SE 13.1.

Results

There were 3337 sore throat illnesses and 735315 days at risk, amongst 7360 participants. Using a Poisson model and weighting the sample to reflect the local age-sex structure and sampling method I calculated the incidence of moderate-severe sore throat episodes as 4.22(95% CI: 3.86-4.61) per 1000-person-days. The median duration of sore throat illness was 7 days (interquartile range 4-10 days). There were 2193 sore throat illnesses self-reported in the seasons where help seeking behaviour was collected (May 2009- March 2011). Approximately ten percent of these illnesses resulted in GP consultation (n=215), and of these more than half took antibiotics (n=121).

Results hereafter are described in children (0-15 years, n=1414) and adults (16 years and older, n=5946) separately.

Children

Sore throat

See Table 8 Incidence and predictors of sore throats in children.

The weighted incidence of sore throat in children (n=1414) was 4.23(95%CI 3.65-4.90) episodes per 1000 person days. Univariable Poisson analyses of measured variables on the incidence of sore throat in children are shown in Table 8; participants who were of school age were more likely to report a sore throat compared to those who were of preschool age (Incidence Rate Ratio – IRR - 1.23). Further analysis showed that sore throat was more likely in those living in regions of higher population density (IRR1.34), suffering from chronic

medical issues (IRR1.32), belonging to the least socially deprived groups (IRR1.39), belonging to a family who are good responders (IRR1.44) and in those living in some regions of England (e.g. South West compared to London IRR 1.45). Therefore, univariable analyses suggested there was regional variation in sore throat.

On multilevel multivariable Poisson modelling the following variables were found to be independently related to the risk of reporting a moderate-severe sore throat episode, even after clustering individuals to their General Practice and adjusting for the effects of other significant variables: Child age (e.g. school age compared to preschool age children - adjusted IRR – aIRR - 1.18); living in regions of higher population density (aIRR 1.29), presence of chronic medical issues (aIRR 1.31) and being from a family who are good responders (aIRR 1.26). Once these variables were accounted for there was no statistical association between region and sore throat incidence.

Systematic component of variation (SCV) was calculated as 2.76, suggesting that after accounting for loco-regional population differences between regions there was very little disparity in the incidence of children with sore throats.

	Sore throat episodes	Time at risk (days)	Incidence rate/1000person days	IRR (95% CI)	P-value	Adjusted IRR (95% CI)	P-value
				Univariable analysis		Multivariable analysis	
Males	362	74319	3.88	0.83(0.64-1.08)	0.16	0.90(0.77-1.07)	0.23
Females	338	63598	4.66	1		1	
Non-white ethnicity	17	4979	3.41	0.66(0.41-1.07)	0.09	-	-
White	591	114465	5.16	1		-	
Rural residence	400	83659	4.78	1	<0.001	1	0.01
Urban residence	297	46458	6.39	1.34(1.15-1.55)		1.29(1.07-1.57)	
Vaccinated	88	16002	5.50	1.10(0.88-1.37)	0.40	-	-
Not vaccinated	599	119710	5.00	1		-	
Chronic illness	89	12869	6.92	1.32(1.05-1.64)	0.02	1.31(1.01-1.68)	0.04
Healthy	592	112629	5.26	1		1	
Most deprived: IMD 1	19	5147	3.69	0.72(0.45-1.15)	0.02	-	-
Least deprived: IMD 5	192	37278	5.15	1		-	
0-4 years	155	35798	3.21	1	0.02	1	0.01
5-15 years	545	102120	4.61	1.23(1.03-1.47)		1.18(0.97-1.44)	
Bad responder	185	45265	3.36	1	<0.001	1	0.02
Good responder	515	92653	4.74	1.41(1.01-1.97)		1.26(1.04-1.52)	
2 people in household	13	4229	3.07	0.69(0.38-1.20)	0.19	-	-
3 people in household	113	24949	4.53	1		-	

4 people in household	367	69360	5.29	1.17(0.95-1.44)		-	
5 people in household	158	30900	5.11	1.13(0.89-1.44)		-	
6 people in household	49	8480	5.78	1.28(0.91-1.78)		-	
North	77	14948	5.15	0.92(0.71-1.19)	0.05	0.97(0.66-1.43)	0.72
West Midlands	44	9345	4.71	0.84(0.61-1.57)		0.85(0.55-1.32)	
East and East Midlands	217	40976	5.30	0.95(0.79-1.13)		0.95(0.74-1.21)	
London	37	9542	3.88	0.69(0.49-0.98)		0.81(0.52-1.26)	
South East	58	15402	3.77	0.67(0.51-0.89)		0.69(0.46-1.03)	
South West	267	47705	5.60	1		1	

Table 8 Incidence and predictors of sore throats in children.

This table demonstrates the incidence and risk of self-reporting a severe sore throat infection (moderate-severe sore throat pain on 2 or more days with fever and no cough) amongst key patient variables, amongst the whole paediatric cohort of FluWatch. It shows the results of univariable and multivariable multi-level Poisson analyses that investigate which of these factors are actually related to self-reporting a sore throat infection in children. The final multi-level Poisson model, presented in this table, clustered at level of patient and practice, with exposure variables as random effects, denominator was 125,498 person-days from 1398 children.

Help seeking behaviour

There were 433 sore throat illnesses reported in children during the seasons when help seeking behaviour was monitored. More than one fifth (n=98) resulted in health contacts being initiated, with GP consultation being the most common form of health contact (n=62) – see Table 9 Types of help seeking behaviour in children with sore throat.

Type of health contact	% of all health contacts (n)
GP consultation	63% (62)
GP phone call	15% (15)
Accident and Emergency	5% (5)
Hospital	4% (4)
Other including NHS Direct/pharmacists/Urgent referral centres etc,.	12% (12)

Table 9 Types of help seeking behaviour in children with sore throat.

This table shows what proportion of children who have self-report a sore throat seek help from these healthcare venues, n= 98 health contacts.

Univariable analyses (see Table 10 Rate and predictors of GP consultation in children with self-reported sore throat.) showed that the following factors were related to consultation amongst those with sore throat: Being preschool age compared to be school age (Odds Ratio – OR - 3.33), living in rural regions (OR2.56), increasing days of illness (OR1.11), reporting severe pain compared to moderate pain (OR23.75), reporting severe earache compared to no earache (OR 3.34) and reporting high fever compared to no fever (OR 9.55). There was no regional difference in GP consultation rate of children who had reported a sore throat.

Multivariable analysis (see Table 10 Rate and predictors of GP consultation in children with self-reported sore throat.) showed that being preschool age compared to school age (adjusted odds ratio – aOR - 2.27), living in rural regions (aOR 2.86), increasing days of sore throat illness (aOR 1.16) and reporting a high fever compared to no fever (aOR 5.32) were all related to increasing risk of GP consultation in children who had a sore throat.

Systematic component of variation was calculated as 1.86, suggesting that after accounting for loco-regional population differences between regions there was very little disparity in the rate of GP consultation behaviour between the six areas of England surveyed.

Variable	GP visits	Total sore throats in this category	%	OR 95	p	Adjusted OR 95	P
Chronically ill	10	50	20%	1.63(0.79-3.36)	0.18	-	-
Well	49	376	13%	1		-	
0-4 years	24	87	28%	1	<0.001	1	0.05
5-15 years	38	346	11%	0.31(0.16-0.59)		0.44(0.18-1.03)	
Female	27	208	13%	1	0.59	1	0.52
Male	35	225	16%	1.19(0.63-2.22)		1.31(0.63-2.70)	
Rural	46	240	19%	1	0.01	1	0.02
Urban	16	190	8%	0.39(0.20-0.76)		0.35(0.14-0.87)	
Non white	5	16	31%	3.17(0.88-11.49)	0.08	-	-
White	52	388	13%	1		-	
Duration (days)	62	433	14%	1.11(1.05-1.16)	<0.001	1.16(1.08-1.24)	<0.001
Severe pain	10	15	67%	-	<0.001	-	-
Moderate pain	6	64	9%	0.04(0.01-0.19)		-	
Mild pain	0	13	-	-		-	
Severe earache	12	37	32%	3.34(1.48-7.76)	0.01	-	-
Mild earache	6	47	13%	0.96(0.34-2.72)		-	
No earache	44	330	13%	1		-	
IMD 1 (Most deprived)	0	6	0%	-	0.66	-	-

IMD 5 (Least deprived)	25	156	16%	1		-	
No fever	11	162	7%	1		1	
Mild fever	25	115	22%	3.99(1.72-9.28)	<0.001	3.47(1.40-8.61)	0.004
High fever	16	39	41%	9.55(3.46-26.38)		5.32(1.69-16.75)	
Good reporters	54	370	15%	1	0.96	-	-
Bad reporters	8	63	13%	1.03(0.39-2.63)		-	
North England	4	42	10%	0.71(0.22-2.30)	0.21	1.17(0.22-6.26)	0.67
West Midlands	4	23	17%	1.37(0.35-5.33)		1.17(0.35-3.84)	
East and East Midlands	26	183	14%	1.10(0.53-2.31)		1.84(0.68-5.00)	
London	8	23	35%	3.48(1.18-10.27)		2.72(0.69-10.68)	
South East	3	31	10%	0.70(0.14-3.59)		2.14(0.29-15.89)	
South West	17	131	13%	1		1	

Table 10 Rate and predictors of GP consultation in children with self-reported sore throat.

This table demonstrates the rate and risk factors of GP consultation in those with those who have already reported a sore throat illness (from start of symptoms to 21days following resolution) in children. It shows the results of univariable and multivariable multi-level logistic analyses that investigate which of these factors are actually related to self-reporting a sore throat infection in children.Multi-level logistic regression clustered at the level of patient and practice, with exposure variables added as random effects, using denominator of 318 sore throat infections.

Adults

Sore throat

See Table 11 Incidence and predictors of sore throats in adult.

The weighted incidence of sore throat in adults (16 years and older, n=5946) was 4.22(95%CI 3.85-4.63) episodes per 1000 person days. Univariable Poisson analyses of measured variables on the incidence of sore throat in adults are seen in Table 11 Incidence and predictors of sore throats in adult.; participants who were 25-44 years old were more likely to report a sore throat compared to 16-24 year olds (OR1.29). Further analyses showed that sore throat was more likely in females (IRR 1.33), those not suffering chronic illness (IRR 1.16), living in households of 4 people compared to households of 2 people (IRR 1.24), being a non-smoker (OR 1.53), being from a family of good reporters (IRR1.37) and those who lived in some regions (e.g. London compared to those who lived in the south west of England (IR 1.27)). Univariable analysis suggests there is regional variation in sore throat incidence.

On multilevel multivariable Poisson modelling (Table 11 Incidence and predictors of sore throats in adult.) the following variables were found to be independently related to the risk of reporting a moderate-severe sore throat episode, even after clustering individuals to their General Practice and adjusting for the effects of other significant variables: Female gender (adjusted OR 1.33); non-white ethnicity (aOR 1.33), being vaccinated against influenza that season (aOR 1.19), being 25-44 years compared to 16-24 years old (aOR 1.23), living in a household of 4 people instead of 2 people (aOR 1.05), non-smokers (aOR 1.54) and being from a family of good reporters (aOR 1.18). Once these variables were accounted for there was no statistical association of region on sore throat incidence.

Systematic component of variation was calculated as 1.2, suggesting that after accounting for measured sociodemographic factors between regions there was very little disparity in the incidence of adults with sore throats between the six areas of England surveyed.

	Sore throat episodes	Time at risk (days)	Incidence rate/1000person days	IRR	P-value	IRR	P-value
				Univariable analysis		Multivariable analysis	
Males	1053	281682	3.72	0.66(0.45-0.97)	<0.001	0.76(0.70-0.83)	<0.001
Females	1584	315715	4.61	1		1	
Non-white ethnicity	82	20603	3.67	0.85(0.67-1.07)	0.17	0.75(0.60-0.95)	0.02
White	2330	497036	4.53	1		1	
Rural residence	1565	342951	4.31	1	0.54	-	-
Urban residence	1061	226526	4.83	1.03(0.94-1.12)		-	
Vaccinated	704	158571	4.39	1.01(0.92-1.10)	0.90	1.19(1.07-1.31)	0.001
Not vaccinated	1862	422014	4.16	1		1	
Chronic illness	447	108435	3.92	0.86(0.77-0.97)	0.01	-	-
Healthy	2133	446987	4.62	1		-	
Most deprived: IMD 1	100	19763	4.63	1.05(0.84-1.31)	0.14	-	-
Least deprived: IMD 5	787	163003	4.58	1		-	
16-24 years	187	42652	4.27	1	<0.001	1	<0.001
25-44 years	755	133926	5.48	1.20.9(1.08-1.53)		1.22(1.02-1.46)	
45-64 years	1222	271103	4.21	1.03(0.87-1.22)		0.93(0.77-1.11)	

65 years and over	473	149716	2.98	0.72(0.58-0.87)		0.62(0.50-0.77)	
1 person in household	180	42891	3.97	1.02(0.85-1.21)		1.08(0.90-1.29)	
2 people in household	1306	316360	4.05	1		1	
3 people in household	430	94133	4.44	1.11(0.99-1.24)	<0.001	1.01(0.88-1.15)	0.03
4 people in household	538	105037	4.98	1.24(1.11-1.39)		1.05(0.92-1.20)	
5 people in household	164	32312	4.79	1.23(1.03-1.46)		1.04(0.86-1.27)	
6 people in household	19	6664	3.05	0.69(0.43-1.12)		0.41(0.24-0.71)	
Non smoker	2498	550051	4.54	1	<0.001	1	<0.001
Smoker	139	47346	2.94	0.65(0.53-0.78)		0.60(0.50-0.73)	
Bad reporter	552	159396	3.46	1	<0.001	1	<0.001
Good reporter	2085	438001	4.76	1.37(1.24-1.52)		1.18(0.91-1.53)	
North	246	65352	3.88	0.90(0.77-1.05)		0.87(0.71-1.05)	
West Midlands	170	39403	3.96	1.03(0.86-1.23)		1.01(0.81-1.26)	
East and East Midlands	801	166131	4.44	1.15(1.04-1.28)	0.002	1.10(0.96-1.26)	0.07
London	180	34018	4.69	1.27(1.06-1.51)		1.26(1.03-1.55)	
South East	366	83467	4.29	1.05(0.92-1.20)		0.97(0.80-1.16)	
South West	874	209026	4.21	1		1	

Table 11 Incidence and predictors of sore throats in adult.

This table demonstrates the incidence and risk of self-reporting a severe sore throat infection (moderate-severe sore throat pain on 2 or more days with fever and no cough) amongst key patient variables, amongst the whole adult cohort of FluWatch. It shows the results of univariable and multivariable Poisson analyses that investigate which of these factors are actually related to self reporting a sore throat infection in children. The final multi-level Poisson model, presented in this table, clustered at level of patient and practice, with all exposure variables as random effects, denominator was 580,585 person-days from 5891 adults..

Help seeking behaviour

There were 1760 sore throat illnesses reported in adults during seasons when help seeking behaviour was monitored. Twelve percent (n=212) resulted in help seeking behaviour, with most adults with sore throat choosing to see their GP (72%) see Table 12 Types of help seeking behaviour in adults with sore throat.

Type of health contact	% of all health contacts (n)
GP consultation	72% (153)
GP phone call	20% (42)
Accident and Emergency	3% (7)
Hospital	2% (5)
Other including NHS Direct/pharmacists/Urgent referral centres etc,.	2% (5)

Table 12 Types of help seeking behaviour in adults with sore throat.

This table shows what proportion of adults who have self-report a sore throat seek help from these healthcare venues, n = 212 health contacts.

Univariable analyses (see Table 13 Rates and predictors of adult GP consultation for sore throat.) showed the following variables were related to the risk of GP consultation during a sore throat illness: Female gender (OR 1.52), chronic medical issues (OR 2.06), increasing days of illness (OR1.11) and severe earache compared to no earache (OR 5.22). In addition, reduced health related quality of life, measured either using the EQ5D-3L index score (OR 0.98) or its five domains (mobility, self-care, usual activities, pain and anxiety – see Table 13 Rates and predictors of adult GP consultation for sore throat. for full results) were all related to increased risk of GP consultation. Adults in the North of England had more than 3.5 times the odds of GP consultation for sore throat compared to adults in the South East of England.

Multivariable analysis (see Table 13 Rates and predictors of adult GP consultation for sore throat.) showed that the only predictor of GP consultation for sore throat in adults was reduction in health-related quality of life, as measured by the composite EQ5D index score (aOR 0.98). After accounting for health-related quality of life no other variables were significantly related to GP consultation, including region of England.

Systematic component of variation was calculated as 4.5, suggesting that after accounting for differences in measured sociodemographics between regions, there was very little disparity in the rate of GP consultation behaviour between the six areas of England surveyed

Variable	GP visits	Total sore throats in this category	%	OR	p	Adjusted OR	P
Chronically ill	41	278	15%	2.06(1.34-3.15)	0.001	-	
Well	111	1460	8%	1		-	
16-24 years	13	106	12%	1	0.38	1	0.41
25-44 years	35	500	7%	0.55(0.27-1.13)		0.29(0.05-1.77)	
45-64 years	76	820	9%	0.74(0.37-1.46)		0.66(0.12-3.53)	
>65 years	29	334	9%	0.69(0.33-1.45)		0.55(0.09-3.40)	
Female	106	1061	9%	1	0.04	1	0.18
Male	47	699	7%	0.66(0.45-0.97)		0.60(0.29-1.26)	
Rural	89	1035	9%	1	0.84	-	-
Urban	64	714	9%	1.04(0.72-1.50)		-	
Non white	5	60	9%	0.98(0.35-2.78)	0.98	-	-
White	143	1640	9%	1		-	
Duration (days)	153	1760	9%	1.11(1.09-1.14)	<0.001	-	-
EQ5D Index score	153	1760	9%	0.98(0.97-0.99)	<0.001	0.98(0.97-0.99)	<0.001
No problems with mobility	28	296	9%	1	0.0 4	-	-
Some problems with mobility	10	51	20%	2.32(1.04-5.16)		-	-
Confined to bed	12	32	38%	5.7(2.29-14.21)		-	-
No problems with self care	39	356	11%	1	0.01	-	-

Some problems washing and dressing	5	15	33%	4.04(1.25-13.07)		-	-
Unable to wash/dress	6	8	75%	24.23(4.11-142.88)		-	-
No problems with usual activities	9	175	5%	1		-	-
Some problems with usual activities	25	155	16%	3.55(1.45-8.73)	0.001	-	-
Unable to perform usual activities	16	49	33%	8.89(1.45-8.73)		-	-
Severe pain	12	37	32%	8.52(1.90-38.18)		-	
Moderate pain	34	267	13%	2.61(0.75-9.07)	0.01	-	-
Mild pain	4	75	5%	1		-	
No anxiety	29	286	10%	1		-	-
Moderate anxiety	19	85	22%	2.63(1.33-5.20)	0.02	-	-
Extreme Anxiety	2	8	25%	2.95(0.32-26.92)		-	-
Severe earache	34	131	26%	5.22(3.21-8.48)		-	
Mild earache	33	209	16%	2.78(1.78-4.36)	<0.001	-	-
No earache	86	1354	6%	1		-	
IMD 1 (Most deprived)	6	53	11%	1.28(0.54-3.01)		-	
IMD 2	10	142	7%	0.72(0.36-1.44)	0.37	-	-
IMD 3	48	479	10%	1.03(0.66-1.61)		-	
IMD 4	34	501	7%	0.67(0.41-1.12)		-	

IMD 5 (Least deprived)	55	574	10%	1		-	
Good reporters	142	1612	9%	0.86(0.45-1.69)	0.66	-	-
Bad reporters	11	148	7%	1		-	
North England	26	172	15%	1	0.05	1	0.20
West Midlands	9	93	10%	0.62(0.28-1.37)		0.73(0.15-3.57)	
East and East Midlands	55	669	8%	0.51(0.30-0.87)		0.85(0.29-2.53)	
London	9	132	7%	0.43(0.18-1.01)		1.52(0.39-5.88)	
South East	9	192	5%	0.28(0.12-0.64)		0.11(0.01-0.94)	
South West	45	502	9%	0.56(0.31-0.99)		0.45(0.13-1.56)	

Table 13 Rates and predictors of adult GP consultation for sore throat.

This table demonstrates the rate and risk factors of GP consultation in those with those who have already reported a sore throat illness (from start of symptoms to 21days following resolution) in children. It shows the results of univariable and multivariable multi-level logistic analyses that investigate which of these factors are actually related to self-reporting a sore throat infection in children. Multi-level logistic regression clustered at the level of patient and practice, with exposure variables added as random effects, using denominator of 1760 sore throat infections.

.

Discussion

Summary

This study demonstrated that the incidence of moderate-severe sore throat episodes, weighted to represent the national population, was 4.22/1000 person years in England. Univariable analyses showed that there was a statistically different incidence of sore throat between six regions of England surveyed for both children and adults. Multi-level multivariable models showed that age, chronic ill health and population density were predictors of sore throat episodes in children, whereas age, gender, ethnicity, smoking status, household size and influenza vaccination status were predictors of sore throats in adults. After accounting for loco-regional population characteristics, there was no difference in the incidence of sore throat in adults or children between regions, as measured by the systematic component of variation or multilevel multivariable models.

Fourteen percent of all sore throat episodes resulted in help seeking behaviour, most of which were GP consultations (10% of all sore throat episodes). Univariable analyses showed that there was a statistically different rate of GP consultation in both children and adults, separately, between the 6 regions of England surveyed. Multilevel multivariable models showed that age, population density, duration of sore throat and presence of fever were predictors of GP consultation in children, whereas reduction in health-related quality of life was the only predictor of GP consultation in adults. After accounting for loco-regional population characteristics there was no difference in the rate of GP consultation for sore throat between regions, as measured by the systematic component of variation or multilevel multivariable models.

Whilst this study showed that there is substantial regional variation in the incidence and GP consultation for sore throat, these disparities appear to be related to differences in population characteristics between regions. It is plausible that these variations in disease occurrence and consultation behaviour may contribute to regional variations in tonsillectomy rates. Strategies that aim to reduce regional variation in tonsillectomy rates may consider educating the public about appropriate sore throat management. That is sore throat infections are self-limiting and can be safely managed with simple rest and analgesia. Only

patients who have red-flag symptoms should seek help from their GP (e.g. Lock jaw, neck swelling, difficulty breathing etc.,)(164-166).

Strengths and weaknesses

This is the largest population-based survey of sore throat to date, weighted to represent the national population. In addition, it is the only study of regional patterns of sore throat incidence to date. The prospective nature of data collection, through daily health diaries and weekly telephone calls, reduced recall bias inherent to retrospective interview studies. This survey method also allowed us to reduce our missing data (14% missing weekly status reports). Sensitivity analyses of different ways of accounting for our missing data showed no change in our conclusions. In contrast to electronic health care record studies, I could accurately assess the role of sore throat severity and associated symptoms in relation to help seeking behaviour. This study has two limitations. Firstly, very young children, residents of Northern England and those of lowest socio-economic status were under-represented in the study population. Therefore, the survey was weighted to allow the incidence to be more representative of local and national populations. Secondly, one year of data collection was conducted in a pandemic influenza outbreak year when there was considerable media coverage, which may have increased symptom vigilance and increased consultation behaviour. The third weakness is that variation can occur at a number of sublevels, however only variation at the level of GPs and 6 large regions was able to be analysed due to a lack of data availability. In the future analyses should attempt to study smaller scale variations.

Relation to published literature

Sore throat

The only other prospective population based study in England was undertaken in Lambeth in 1974 on 198 women, aged 20-44, who were asked to keep a prospective health diary for 28-days each. During the observation period 90 sore throat episodes were reported (annual sore throat incidence of 5.9 (95% CI 4.7-7.3) sore throat episodes per person-year) with 33 subsequent GP consultations (37% (95% CI 25-51%) consultation rate. However, since this was a small study in a select population it is difficult to draw meaningful comparisons. Our results add to this growing body of evidence from general population studies of respiratory

infections (103,104,108,167) that the majority of sore throats, are managed safely in the community.

Help seeking behaviour

I found that GP consultation for sore throat was predicted by young age. Studies of all respiratory infections also confirm that young age is a major driver for primary care use in UK (69) with qualitative studies showing that parents' decision to bring their children to the GP is influenced by perceived threat, disease severity, the perceived benefits of consulting, and an expectation of assessment, information, advice or treatment (69-72,168). Our study also found duration of sore throat and presence of fever were related to GP consultation in children. Explanations for these results can be offered from qualitative research into people with acute sore throat(169) and respiratory tract infections (167) showing that people most commonly seek help from their GP for perceived symptom severity and non-resolution of symptoms.

Whilst GP consultation is only one type of help seeking behaviour, it is pertinent to our overall study of tonsillectomy, since a patient is almost exclusively referred for a tonsillectomy through primary care. Whilst several models of help seeking behaviour have been described in the literature (170,171), a framework of outcome and health behaviour has been proposed (172) that encompasses several social-cognition models. The Shaw model describes that the experience of symptoms and the subsequent health behaviour are based on an individual's appraisal of the symptom as being a health threat, followed by an assessment of the severity of the health threat and the formation of behavioural intentions. Appraisal of symptom severity is dependent on (a) impact of symptoms on routines of life; (b) an appraisal of coping resources mediated by personality, locus of control, social support, preferred coping styles, expectations and self-efficacy; and (c) individual differences such as age, gender, social status and ethnicity (which reflect wider cultural influences). Since a sore throat is a common and self-limiting illness, it is reasonable to think that help seeking was initiated when the perceived symptom severity was beyond the patient's personal coping strategies. Our results seem to follow this model of help –seeking behaviour.

In children, our study showed that GP consultation for sore throat was predicted by very young age. Epidemiological studies confirm that very young age is a major driver for primary care use in UK(69). Qualitative studies have shown that very young children are perceived as vulnerable and there is more frequent delegation of care(70-72). In addition, I found markers of disease severity such as fever and increasing duration were also strong predictors of GP consultation in children. Qualitative studies have shown that parents' decision to bring their children to the GP is influenced by perceived threat, disease severity, the perceived benefits of consulting, and an expectation of assessment, information, advice or treatment (69-72,168). Therefore, it is likely that parents in our study chose to consult their GP when the perceived threat of the sore throat illness was beyond their coping strategies. Children residing in rural areas were more likely to consult their GP than children with similar disease severity in urban areas. Since there was a lower incidence of sore throat in rural regions, parents may have been less exposed to these illnesses and perceive them with more threat. Alternatively, despite the potential need to travel further in rural areas there could have been easier access to services in rural areas that aren't burdened with over population, which in turn may have lowered the threshold for consultation. Further qualitative studies may help answer this question.

In adults, our study showed that GP consultation for sore throat was predicted by lower self-reported health status (as measured by the EQ-5D 3L). Low health status, may have meant that respondents had difficulty coping with mobility, usual activities, self-care, or reported high pain, or anxiety. This index score, therefore, may be associated with a high perception of threat or an exhaustion of coping strategies. Hence according to the framework of help seeking behaviour EQ5D may be a good surrogate measure of perceived symptom severity. A sore throat specific health related quality of life score may improve our understanding of this relationship.

Implications and future work

This is the first study of regional disparities in sore throat disease burden. It shows that there is considerable variation in the incidence and help seeking behaviour for sore throat between six regions of England, both for children and adults. However, the variations in disease burden are predicted by variations in local population characteristics such as age, gender, population

density, chronic illnesses, influenza vaccination status and smoking status. After controlling for these characteristics there was no measurable variation in GP consultation rates for sore throats between regions. Unfortunately, studies of geographical practice variations rarely have the granularity in their data to account for such characteristics and therefore, may inadvertently overestimate the unwarranted variation in treatment rates. Whilst I have shown in this data that there is no variation in GP consultation rates between regions, after accounting for local population characteristics, the overall number of participants who consulted a GP is low and further work needs to be undertaken to corroborate our findings on a larger population sample.

CHAPTER FOUR: Is there variation in the incidence of recurring sore throat and the rates of tonsillectomy between regions of England?

Chapter synopsis

In this chapter, I have summarised the results of a large retrospective electronic cohort study of sore throat following health care contact: That is of sore throat consultations in primary care, recurring sore throat as defined by consultations in primary care, and tonsillectomy rates in secondary care. I was able to undertake these investigations following linkage of primary and secondary care databases. These investigations demonstrated that whilst there was substantial variation in the rates of sore throat, recurring sore throat and tonsillectomy between the 10 regions of England, once regional population characteristics were accounted for, variation reduced considerably, to become statistically non-significant.

Introduction

Previous epidemiological studies have shown considerable variation in tonsillectomy rates, both between countries (159)) and regions within a country (16,17,19,20). This observation has since been noted for many other surgical conditions(17,24-26) (27) (28-30) (31,32)) (33-35) and considerable effort has gone into reducing regional surgical rate variation, broadly described within the evidence based medicine movement(47). Unfortunately, and specific to tonsillectomy, the regional surgical rate variation persists (20). There is now a growing body of evidence, summarised and categorised in chapter one, that comes from public health and decision analysis studies suggesting regional healthcare variations are a direct response to the regional variations in healthcare 'need'(61) (62) (63) (64,65), (66) (67) (68,69) (70-72) (77) (78-80) (81-84) (85-88) (89) (90) (91). However, this information has not been described in relation to regional tonsillectomy rate variations. This may partly relate to the fact that records for tonsillitis, and recurring tonsillitis diagnoses are kept in primary care databases where the condition is most frequently managed. In contrast, tonsillectomy is undertaken, and recorded, solely in secondary health care setting. Historically, there has not been any linkage of health care information between these settings on a large enough scale to describe regional surgical rate variations. This means relating regional tonsillectomy rate variations to regional incidence of recurring tonsillitis (key component of 'need' for tonsillectomy) has never been undertaken. In fact, the previous chapter has shown that there is considerable unadjusted regional variation in the occurrence of sore throat in the community and rate of help seeking behaviour – implying a potential role for regional variation in 'need'. We do not yet know how the disparities in community sore throat and help seeking behaviour are related to tonsillectomy rate variation.

Aims

Therefore, in this chapter I aimed to investigate regional variations in the primary care consultations for sore throat and recurring sore throat and their relationship to geographical disparities in tonsillectomy rates, as measured in the secondary care setting.

Objectives

I attempted to achieve my aims by answering 3 main questions, separately for sore throat, recurring sore throat and tonsillectomy:

A. Is there regional variation in sore throat/recurring sore throat/tonsillectomy rates?
B. What are the predictors of sore throat/recurring sore throat/tonsillectomy rates?
C. How much regional disparity persists once these predictors of sore throat/recurring sore throat/tonsillectomy are accounted for?

Methodology

The study used the CALIBER database, which is broadly representative of the national cohort and as such have comparable characteristics in terms of age, gender and socioeconomic status(113,173).

I divided the CALIBER dataset into children (preschool-0-4 and school 5-15 years old) and adults (young adults 16-24 and early adulthood 25-44 years) as there was evidence to suggest that the epidemiology and management of sore throat was different in these populations. Finally, adults over the age of 44 were rarely at risk of recurring sore throat or tonsillectomy for recurring sore throat, therefore, I removed them from analyses as I felt that would artificially and inappropriately inflate the denominator (see Table 14 Sore throat, recurring sore throat and tonsillectomies by age group.).

For a full description of variable definitions, please refer to Chapter 2.

Age category (years)	Total person-years at risk	Patients with recurring sore throat *n* (% of those who had sore throat)	Total number of patients who had tonsillectomy for recurring tonsillitis
0-4	347,734	4871 (6%)	1108
5-15	2,253,956	13385 (9%)	3170
16-24	2,360,050	12,726 (7%)	1675
25-44	5,347,471	10,381 (4%)	981
>44	$1.06_{10}7$	4080(2%)	106
Total number	$2.10_{10}7$	45,443	7040

Table 14 Sore throat, recurring sore throat and tonsillectomies by age group.

This table demonstrates that only 2% of adults over the age of 44 go from having one sore throat to being defined as those with recurring sore throat (3 consultations in less than 1 year). They also account for less than 2% of all the tonsillectomies done in those with recurring tonsillitis, whilst accounting for nearly half the denominator time ($2.10_{10}7$ person years)

Modelling

Similarly to the study described in Chapter 3, since participants could visit their GP more than once during their time at risk, logistic models were less appropriate. General linear models (GLM) were used to quantify the rate and predictors of sore throat, as they allowed repeated counts and could account for disparities in the time at risk. GLM models were also used to calculate the incidence of recurring sore throat and tonsillectomy rates in the cohort as they accounted for variations in time at risk. I assessed how different models (Poisson, Zero-inflated Poisson and Negative binomial) predicted dispersion of the data and the participants that never reported sore throats. Poisson modelled our dataset best and was therefore used for all analyses.

Outcome variables

For the first set of analyses the outcome variable was GP sore throat consultation. For more detailed information on how this variable was defined please refer to chapter 2. In summary, it was defined as any GP consultation that was coded with any one or more of the 111 sore throat READ codes defined in Appendix U. GP sore throat consultation was used as a surrogate marker for GP tonsillitis consultation, with the assumption that the proportion of sore throat consultations that are actually tonsillitis will not vary from region to region. The entire CALIBER cohort was used as study denominator, with children (denominator 3.0million person-years at risk) and adults (denominator 7.7million person-years at risk) analysed separately, and time at risk censored for 21 days around a sore throat consultation, as it was not felt that patients were at risk of a new sore throat infection. As participants could attend their GP more than once with a sore throat I clustered data to the level of the participant in these analyses. In addition, further clustering of patients to general practices, and general practices to regions was found to better describe the data on likelihood ratio test against a single level model. Therefore, I used multi-level (patient, general practice and regional level) models to evaluate clinical and sociodemographic determinants of sore throat GP consultation.

In the second set of analyses the outcome variable was recurring sore throat. This was defined as any patient who had 3 or more GP sore throat consultations (defined above) in 12months or less. The entire CALIBER cohort was used as study denominator, with time at risk censored

when a patient becomes defined as having recurring sore throat. Children (denominator 2.9million person-years at risk) and adults (denominator 7.6million person-years at risk) were analysed separately.

In the final set of analyses, tonsillectomy was the outcome variable. Since not all tonsillectomies are done for recurring tonsillitis, and hospitals do not use ICD-10 codes for diagnosis reliably, the outcome was defined as any patient with recurring GP sore throat consultations (defined above) who went on to receive an OPCS code for tonsillectomy.

The entire CALIBER cohort was used as study denominator, with children (denominator 2.6million person-years at risk) and adults (denominator 7.7million person-years at risk) analysed separately, and time-at-risk censored once a patient received a tonsillectomy.

Since patients were not able to have more than one diagnosis of recurring sore throat or one tonsillectomy, analyses that had these variables as the main outcome were clustered at the level of general practice and region only. Exposure Variables Definitions of exposure variables are provided in chapter 2. In summary, I used the following variables: Ethnicity; Geographical region (10 geographically based Strategic Health Authorities that manage local health care); Social deprivation (categorised into national quintiles based on the index of multiple deprivation; Respiratory disease in children as asthma an adult having been given the diagnosis of asthma or chronic obstructive pulmonary disease; Obesity (body mass index greater than 29, recorded at least once during their window of observation); Obstructive sleep apnoea; HIV status; Eating disorder; Alcohol consumption (Non-drinkers, Mild-moderate drinker, Heavy drinkers); Smoking status (non-, ex- and current smoker); Diabetes; and Hypertension.

Exposure variables that were associated with the outcome with a *p* value of <=0.1, on univariable analysis, were considered for inclusion in multivariable regression models with an *a priori* decision to include age, gender and ethnicity as patient characteristics, and social deprivation (since IMD is calculated by postcode), regardless of the association. Additional variables were included initially starting with variables that had the most number of plausible mechanisms of action and then adding those with fewer plausible mechanisms. Exposure variables were added sequentially and hierarchically, as random effects to the appropriate

level, if they improved the model fit. Akaike's information criteria (AIC) (160) for sequential models were noted and the probability that each model could reduce the information loss as compared to the model with lowest AIC was calculated. The model was retained if it had a high probability of reducing information loss. I undertook tests for interaction between variables if both variables were independently related to the outcome and had a biologically plausible interdependent relationship to outcome. Once a full multi-level multivariable model had been created, region was added to the model to see if geography was still important after accounting for population and disease level predictors of sore throat, recurring sore throat and tonsillectomy rates.

Models developed to predict the GP sore throat consultation rates, incidence of recurring sore throat and tonsillectomy rates were used to calculate the systematic component of variation (SCV). The SCV is described in chapter 3. The SCV was calculated using all variables found to be significant in multi-level multi-model testing.

All statistical analyses were undertaken in Stata SE 13.1.

Results

Our cohort consisted of 3,560,864 patients making up more than 21 million person-years of follow up. Within our cohort there were 1,440,002 sore throat consultations (68.0/1000person-years) and 16,618 tonsillectomies (7.9 tonsillectomies for 10,000person years of follow up), both of which are comparable to the expected levels(2,174). The sore throat consultation rate observes in CPRD was around half the expected sore throat consultation rate as predicted by the FluWatch study (around 154 consultations/1000 patient year). However, FluWatch did not have full person years but winter seasons only.

A quarter of patients in our cohort – all ages - (*n=861,600*) saw their GP at least once for a sore throat (mean: 1.67 consultations per patient seeing their GP at least once with sore throat). Five percent of those who had consulted for a sore throat once (*n=45,443*) went on to have recurring sore throat (3 sore throat consultations within a 12-month period). Sixteen percent (*n=7040*) of patients with recurring sore throat went on to have a tonsillectomy. Table 14 shows the rate of sore throat, recurring sore throat (3 sore throat consultations within 12

months) and tonsillectomy rate by age category. From this point forward all analyses will be described as either for children (0-15 years old) or for at risk adults (16-44 years old).

Children

General

There were 410,477 sore throat consultations over 2,456,253 person years of observation giving a sore throat consultation rate of 167/1000person years (95% CI 166-168) (comparable to predicted paediatric consultation rate based on FluWatch 200/1000 person years). 18,256 children suffered with recurring sore throat, giving an annual incidence of 6.9/1000 person-years (95%CI 6.76-6.96). Whilst 7849 children were recorded as having a tonsillectomy for recurring tonsillitis, only 3361 met our definition of recurring sore throat (3 sore throat consultations in 12 months).

Sore throat

See Table 15 Multi-variable models of sore throat consultation in children.

Univariable analysis showed a small amount of variation in the rate of sore throat consultation by geographical region (Residents of East Midlands had 33% more chance of having a sore throat consultation compared to residents from the North East). Please see Appendix D for univariable analyses. Running the Systematic Component Variation (SCV) model without adjustment of population characteristics showed the SCV value of 4.2.

Multi-level multivariable Poisson modelling showed the following factors were more likely to be related to GP sore throat consultation in children: Very young age (0-4 years: adjusted incidence rate ratio – aIRR – 1.81), female gender (aIRR 1.24), ethnicity (e.g. white ethnicity compared to Black African – aIRR 1.62), lower deprivation score (least deprived compared to most deprived aIRR 1.11), chronic respiratory illness (aIRR1.79), obesity (aIRR 2.00), eating disorders (aIRR 1.88), and HIV (aIRR 1.30).

After controlling for all the above factors there was very slightly less variation in sore throat incidence between the 10 regions of England (Residents of East Midlands now had 32% more chance of a sore throat consultation compared to Residents of the North East). However,

there was more unaccounted variation between general practices within regions than between the regions themselves. The multilevel model revealed that there was 5 times more unaccounted variation at the general practice level compared to the region level (variance 0.003 vs 0.015, respectively). The above multilevel multivariable model was used to calculate a new systematic component of variation (adjusted SCV). The adjusted SCV score of 1.2 showed that much of the regional disparity in sore throat can be predicted by regional population characteristics.

Characteristic	Multivariable analysis	
	Adjusted IRR	P
Gender		
Male	1	
Female	1.24(1.23-1.26)	<0.001
Age category		
0-4 years	1	
5-15 years	0.55(0.55-0.56)	<0.001
Ethnic origin		
White British	1	<0.001
Indian	0.87(0.82-0.93)	
Black African	0.63(0.59-0.68)	
Black Caribbean	0.68(0.62-0.74)	
Black other	0.69(0.63-0.75)	
Bangladeshi	1.01(0.90-1.13)	
Pakistani	0.96(0.91-1.02)	
Other Asian	0.82(0.75-0.89)	
Chinese	0.72(0.62-0.82)	
Mixed	0.67(0.63-0.71)	
Other	0.85(0.81-0.89)	
Unknown	0.92(0.90-0.93)	
Social Deprivation		
Least deprived	1	
Most deprived	0.94(0.92-0.96)	<0.001
Respiratory illness		
Absent	1	
Present	1.79(1.76-1.81)	<0.001

Obstructive sleep apnoea		
Absent	1	<0.001
Present	1.79(1.76-1.81)	
Obesity		
Not coded	1	
Obese	2.00(1.93-2.08)	<0.001
HIV status		
HIV negative	1	
HIV positive	1.30(1.17-1.45)	<0.001
Eating disorder		
Absent	1	
Eating disorder	1.88(1.76-2.01)	<0.001
Practice region		
North East	1	
North West	1.23(1.15-1.32)	
Yorkshire & The Humber	1.14(1.06-1.23)	
East Midlands	1.32(1.22-1.42)	
West Midlands	1.37(1.27-1.47)	
East of England	1.24(1.16-1.34)	
South West	1.07(0.99-1.15)	
South Central	1.07(0.99-1.15)	
London	1.17(1.09-1.26)	
South East Coast	1.26(1.017-1.35)	<0.001

Table 15 Multi-variable models of sore throat consultation in children.

This table shows the risk factors associated with sore throat consultation, amongst all children in the CALIBER cohort (denominator=3.0million person years). Multivariable model is based on the Poisson distribution. Multivariable models are multi-level and include random effects.

Recurring sore throat

See Table 16 Multi-variable models of recurring sore throat in children.

Univariable analysis showed a moderate amount of variation in the incidence of recurring sore throat rates by geographical region (Residents of East Midlands had 86% more chance of having recurring sore throat compared to residents from the North East). Running the SCV model without adjustment of population characteristics showed the SCV value of 4.8.

Multilevel multivariable Poisson modelling showed the following factors were more likely to be related to children defined as suffering from recurring sore throat: Very young age (0-4 compared to 5-15 years: aIRR 1.56), female gender (aIRR 1.40), ethnicity (e.g. white ethnicity compared to Black African – aIRR 2.38), lower deprivation score (least deprived compared to most deprived aIRR 1.12), chronic respiratory illness (aIRR2.03), obstructive sleep apnoea (aIRR 2.10), obesity (aIRR 2.00) and eating disorders (aIRR2.30).

After controlling for all the above factors there was less variation in recurring sore throat rates between the 10 regions of England (Residents of East Midlands now had 65% increased risk of a recurring sore throat compared to Residents of the North East). There was more unaccounted variation between general practices within regions than between the regions themselves. The multilevel model revealed that there was four times more unaccounted variance at the level of the general practice compared to the level of the region (0.54 vs 0.12, respectively). The above multilevel multivariable model was used to calculate the systematic component of variation. The adjusted SCV score of 1.8 showed that much of the regional disparity in recurrent sore throat can be predicted by regional population characteristics.

Risk Factors	Multivariable model	
	Adjusted IRR	P
Gender		
Male	1	<0.001
Female	1.40(1.36-1.45)	
Age category		
0-4 years	1	<0.001
5-15 years	0.64(0.62-0.67)	
Ethnic origin		
White British	1	<0.001
Indian	0.67(0.55- 0.82)	
Black African	0.42(0.33- 0.54)	
Black Caribbean	0.50(0.38- 0.67)	
Black other	0.61(0.47- 0.81)	
Bangladeshi	1.12(0.84-1.48)	
Pakistani	0.90(0.76- 1.06)	
Other Asian	0.70(0.54- 0.90)	
Chinese	0.38(0.21- 0.68)	
Mixed	0.45(0.37- 0.55)	
Other	0.79(0.68- 0.92)	
Unknown	0.83(0.80- 0.87)	
Social Deprivation		
Least deprived	1	<0.001
Most deprived	0.89(0.84-.094)	
Respiratory illness		
Absent	1	<0.001
Present	2.03(1.95-2.10)	
Obstructive sleep apnoea		
Absent	1	<0.001
Present	2.10(1.69-2.63)	

Obesity		
Not coded	1	<0.001
Obese	2.31(2.12-2.53)	
Eating disorder		
Absent	1	<0.001
Eating disorder	2.30(1.95-2.70)	
Practice region		
North East	1	<0.001
North West	1.41(1.11- 1.78)	
Yorkshire & The Humber	1.42(1.13-1.79)	
East Midlands	1.65(1.31-2.08)	
West Midlands	1.68(1.33- 2.13)	
East of England	1.49(1.17- 1.88)	
South West	1.11(0.88- 1.41)	
South Central	1.01(0.79- 1.28)	
London	1.29(1.01- 1.63)	
South East Coast	1.49(1.18- 1.89)	

Table 16 Multi-variable models of recurring sore throat in children.

*This table shows the risk factors associated with recurring sore throat consultation, amongst **all** children in the CALIBER cohort (denominator=2.9million person-years). Multivariable models are based on the Poisson distribution. Multivariable models are multi-level and include random and fixed effects.*

Tonsillectomy

See Table 17 Multi-variable models of tonsillectomy in children.

Univariable analysis showed considerable variation in tonsillectomy rates by geographical region (Residents of East Midlands had 290% more chance of a tonsillectomy compared to residents from the North East). Running the SCV model without adjustment of population characteristics showed the SCV value of 8.8.

Multilevel multivariable Poisson modelling of tonsillectomy rates in all children showed the following factors were more likely to be related to receiving a tonsillectomy: School age (5-15 years compared to 0-4 years: aIRR – 5.34), female gender (aIRR 1.10), ethnicity (e.g. white ethnicity compared to Black African – aIRR 1.40), lower deprivation score (least deprived compared to most deprived aIRR 1.47), chronic respiratory illness (aIRR1.50), obstructive sleep apnoea (aIRR 10.99), obesity (aIRR 1.96) and eating disorders (aIRR1.54).

After controlling for all the above factors there was still variation in tonsillectomy rates between the 10 regions of England (Residents of East Midlands had aIRR 3.04 for a tonsillectomy consultation compared to Residents of the North East). The multilevel model revealed that there was more variance unaccounted for at the level of the general practices within regions compared to between regions (variance 0.38 vs 0.23, respectively). The final multilevel multivariable model was used to calculate the systematic component of variation. The SCV score of 2.25 showed that there was little disparity between regional tonsillectomy rates after controlling for regional population characteristics.

Describing this more intuitively Table 18 shows that whilst the numbers of tonsillectomies differ considerably per region the proportion of children with recurring tonsillitis is similar between regions.

Characteristic	Multivariable analysis	
	Adjusted IRR	P
Gender		
Male	1	<0.001
Female	1.10(1.05-1.16)	
Age category		
0-4 years	1	<0.001
5-15 years	5.34(4.64-6.15)	
Ethnic origin		
White British	1	<0.001
Indian	0.86(0.66-1.13)	
Black African	0.71(0.52-0.97)	
Black Caribbean	0.50(0.34-0.74)	
Black other	0.86(0.62-1.20)	
Bangladeshi	0.61(0.35-1.06)	
Pakistani	1.01(0.80-1.28)	
Other Asian	0.94(0.65-1.34)	
Chinese	0.47(0.21-1.04)	
Mixed	0.92(0.73-1.16)	
Other	0.89(0.73-1.08)	
Unknown	0.62(0.58-0.66)	
Social Deprivation		
Least deprived	1	<0.001
Most deprived	0.68(0.62-0.74)	
Respiratory illness		
Absent	1	<0.001
Present	1.50(1.42-1.59)	
Obstructive sleep apnoea		
Absent	1	<0.001

Present	10.99(9.7-12.40)	
Obesity		
Not coded	1	<0.001
Obese	1.96(1.66-2.30)	
HIV status		
HIV negative	-	
HIV positive	-	
Eating disorder		
Absent	1	0.01
Eating disorder	1.54(1.13-2.12)	
Practice region		
North East	1	<0.001
North West	1.92(1.30-2.84)	
Yorkshire & The Humber	1.80(1.20-2.69)	
East Midlands	3.04(2.04-4.54)	
West Midlands	2.01(1.35-2.97)	
East of England	2.27 (1.53-3.36)	
South West	1.72(1.16-2.55)	
South Central	1.77(1.19-2.63)	
London	1.87(1.26-2.78)	
South East Coast	2.75(1.86-4.08)	

Table 17 Multi-variable models of tonsillectomy in children.

*This table shows the risk factors associated with tonsillectomy, amongst **all** children in the CALIBER cohort (denominator=2.6million person-years). Univariable and Multivariable models are based on the Poisson distribution. Multivariable models are multi-level and include random and fixed effects.*

Table 18Table of numbers and proportions of children with recurring sore throat who then receive tonsillectomy, by region

Practice region	Children with recurring sore throat patients	Children receiving tonsillectomy for recurring tonsillitis	Proportion
North East	105	24	23%
North West	3,699	1065	29%
Yorkshire & The Humber	1,026	289	28%
East Midlands	983	358	36%
West Midlands	3,174	747	24%
East of England	2,935	818	28%
South West	2,241	667	30%
South Central	1,884	566	30%
London	2,552	620	24%
South East Coast	2,413	770	32%
Mean	2101	592	28%

Adults

General

There were 766,302 sore throat consultations over 8,268,459 person years of observation giving a sore throat consultation rate of 92.7/1000 person years (95% CI 92.5-92.9). 20,423 adults went on to have recurring sore throat giving an annual incidence of 2.67/1000person years (95%CI 2.80-2.87). Whilst 7894 adults were recorded as having a tonsillectomy for

recurring tonsillitis, only 3568 were in our cohort of adults with recurring sore throat (3 sore throat consultations in 12 months).

Sore throat

See Table 19 Multivariable models of Sore throat in adults 16-44 years old.

Univariable analysis showed a small amount of variation in the rate of sore throat consultation by geographical region (Residents of East Midlands had 24% more chance of having a sore throat consultation compared to residents from the North East). Running the SCV model without adjustment of population characteristics showed the SCV value of 3.4.

Multi-level multivariable Poisson modelling showed the following factors were more likely to be related to GP sore throat consultation in adults (16-44years): Young age (16-24 years: aIRR 1.45), female gender (aIRR 1.75), ethnicity (e.g. white ethnicity compared to Black African – aIRR 1.20), low deprivation score (least deprived compared to most deprived aIRR 1.09), presence of chronic respiratory illness (aIRR1.26), diagnosis of obstructive sleep apnoea (aIRR 1.45), obesity (aIRR 1.45), HIV (aIRR 1.37), being a non-drinker compared to heavy drinker (aIRR 1.54), being an active smoker (aIRR 1.10), being diabetic (aIRR 1.20) and being hypertensive (aIRR1.40).

After controlling for all the above factors there was less variation in sore throat incidence between the 10 regions of England (Residents of East Midlands had 18% greater risk of a sore throat consultation compared to Residents of the North East). However, much of the variation was seen to occur between general practices rather than regions. The multilevel model revealed that there was 4 times more unaccounted variation at the general practice level compared to the region level (variance 0.003 vs 0.012, respectively).

The above multilevel multivariable model was used to calculate the systematic component of variation. The adjusted SCV score of 1.7 showed that much of the regional disparity in sore throat can be predicted by regional population characteristics.

Characteristic	Multivariable analysis	
	Adjusted IRR	P
Gender		
Male	1	<0.001
Female	1.75(1.71-1.79)	
Age category		
16-24 years	1	<0.001
25-44 years	0.51(0.49-0.52)	
Ethnic origin		
White British	1	<0.001
Indian	1.06(0.98-1.14)	
Black African	0.83(0.72-0.95)	
Black Caribbean	0.87(0.79-0.97)	
Black other	0.85(0.75-0.95)	
Bangladeshi	1.12(0.97-1.29)	
Pakistani	1.22(1.13-1.32)	
Other Asian	0.97(0.86-1.10)	
Chinese	0.78(0.68-0.90)	
Mixed	0.80(0.71-0.89)	
Other	0.99(0.93-1.05)	
Unknown	0.96(0.93-0.98)	
Social Deprivation		
Least deprived	1	0.01
Most deprived	1.09(1.00-1.19)	
Respiratory illness		
Absent	1	<0.001
Present	1.26(1.23-1.29)	

Obstructive sleep apnoea		
Absent	1	<0.001
Present	1.45(1.31-1.61)	
Obesity		
Not coded	1	<0.001
Obese	1.45(1.26-1.43)	
HIV status		
HIV negative	1	<0.001
HIV positive	1.37(1.24-1.51)	
Alcohol		
Non-drinker	1	<0.001
Mild-Moderate drinker	1.03(0.99-1.07)	
Heavy drinker	0.65(0.60-0.71)	
Smoking		
Non-Smoker	1	<0.001
Ex-smoker	1.17(1.14-1.21)	
Smoker	1.10(1.08-1.12)	
Diabetes		
No Diabetes coded	1	<0.001
Diabetes coded	1.20(1.15-1.26)	
Hypertension		
No hypertension coded	1	<0.001
Hypertension coded	1.44(1.39-1.50)	
Practice region		
North East	1	0.01
North West	1.13(0.84-1.52)	
Yorkshire & The Humber	1.07(0.77-1.49)	
East Midlands	1.18(0.86-1.61)	
West Midlands	1.23(0.92-1.64)	

East of England	1.14(0.83-1.54)
South West	1.12(0.82-1.52)
South Central	0.98(0.72-1.32)
London	0.97(0.71-1.31)
South East Coast	0.99(0.73-1.36)

Table 19 Multivariable models of Sore throat in adults 16-44 years old.

*This table shows the risk factors associated with recurring sore throat consultation, amongst **all adults 16-44 years old** in the CALIBER cohort 1997-2010, denominator is 7.7 million person-years. Multivariable models are based on the Poisson distribution. Multivariable models are multi-level and include random and fixed effects*

Recurring sore throat

See Table 20 Multi-variable models of recurring sore throat in adults (16-44 years old).

Univariable analysis showed a moderate amount of variation in the incidence of recurring sore throats between geographical regions (Residents of East Midlands had 50% more chance of having recurring sore throat compared to residents from the North East). Running the SCV model without adjustment of population characteristics showed the SCV value of 4.2.

Multilevel multivariable Poisson modelling showed the following factors were more likely to be related to adults (16-44years) defined as suffering from recurring sore throat (3 GP sore throat consultations within 12 months): Young age (16-24 years: aIRR 1.52), female gender (aIRR 1.98), ethnicity (e.g. white ethnicity compared to Black African – aIRR 1.47), chronic respiratory illness (aIRR 1.53), obstructive sleep apnoea (aIRR 1.82), obesity (aIRR 1.55), being a non-drinker compared to a heavy drinker (aIRR 2.08), being an active smoker (aIRR1.12), and being hypertensive (aIRR 1.26).

After controlling for all the above factors there was still variation in the incidence of recurring sore throats between the 10 regions of England (e.g. Residents of East Midlands had a 54% increased risk for recurring sore throat compared to Residents of the North East). However, much of the variation was seen to occur between general practices rather than regions. The multilevel model revealed that there was 3 times more unaccounted variation at the general practice level compared to the region level (variance 0.14 vs 0.45, respectively). The final multilevel multivariable model was used to calculate the systematic component of variation. The adjusted SCV score of 2.1 showed that much of the regional disparity in sore throat can be predicted by regional population characteristics.

Characteristic	Multivariable analysis	
	Adjusted IRR	P
Gender		
Male	1	<0.001
Female	1.98(1.82-2.15)	
Age category		
16-24 years	1	<0.001
25-44 years	0.66(0.61-0.71)	
Ethnic origin		
White British	1	0.004
Indian	0.95(0.74-1.23)	
Black African	0.68(0.44-1.03)	
Black Caribbean	0.61(0.45-0.83)	
Black other	0.78(0.58-1.03)	
Bangladeshi	1.14(0.75-1.73)	
Pakistani	1.18(0.99-1.40)	
Other Asian	1.04(0.74-1.45)	
Chinese	0.69(0.43-1.13)	
Mixed	0.92(0.69-1.24)	
Other	0.82(0.65-1.05)	
Unknown	0.89(0.82-0.96)	
Respiratory illness		
Absent	1	<0.001
Present	1.53(1.44-1.63)	
Obstructive sleep apnoea		
Absent	1	<0.001
Present	1.82(1.34-2.48)	
Obesity		

Not coded	1	<0.001
Obese	1.55(1.40-1.71)	
Alcohol		
Non-drinker	1	<0.001
Mild-Moderate drinker	0.94(0.87-1.01)	
Heavy drinker	0.48(0.40-0.58)	
S**moking**		
Non-Smoker	1	0.003
Ex-smoker	1.06(0.96-1.17)	
Smoker	1.12(1.06-1.18)	
Hypertension		
No hypertension coded	1	<0.001
Hypertension coded	1.26(1.13-1.41)	
Practice region		
North East	1	0.002
North West	1.28(0.87-1.90)	
Yorkshire & The Humber	1.23(0.79-1.92)	
East Midlands	1.54(1.02-2.31)	
West Midlands	1.44(0.98-2.10)	
East of England	1.19(0.80-1.78)	
South West	1.18(0.79-1.78)	
South Central	0.89(0.60-1.32)	
London	1.01(0.67-1.51)	
South East Coast	1.03(0.69-1.54)	

Table 20 Multi-variable models of recurring sore throat in adults (16-44 years old).

*This table shows the risk factors associated with recurring sore throat consultation, amongst **all adults 16-44 years old** in the CALIBER cohort, denominator= 7.6million person-years. Multivariable models are based on the Poisson distribution. Multivariable models are multi-level and include random and fixed effect*

Tonsillectomy

See Table 21 Multi-variable models of tonsillectomy in adult.

Univariable analysis showed a small amount of variation in the rate of tonsillectomies between geographical regions (Residents of East Midlands had 21% more chance of having a tonsillectomy compared to residents from the North East). Running the Systematic Component Variation (SCV) model without adjustment of population characteristics showed the SCV value of 5.3.

Multilevel multivariable Poisson modelling in all adults (16-44 years) within the full CALIBER cohort showed the following factors predicted a tonsillectomy: Young age (16-24 years: aIRR – 6.25), female gender (aIRR 1.39), ethnicity (e.g. white ethnicity compared to Black African – aIRR 2.50), chronic respiratory illness (aIRR 1.18), obstructive sleep apnoea (aIRR 2.59), obesity (aIRR 1.58), being a non-drinker compared to heavy drinker (aIRR 1.64) and being an active smoker (aIRR 1.30).

After controlling for all the above factors there was no variation in tonsillectomy rates between the 10 regions of England. Additionally, much of the variation was seen to occur between general practices rather than regions. The multilevel modelled revealed that there was comparable unaccounted variation between the general practice level and the region level (variance 0.24 vs 0.27, respectively). The above multilevel multivariable model was used to calculate the systematic component of variation. The adjusted SCV score of 2.91 showed that much of the regional disparity in tonsillectomy rates can be predicted by regional characteristics. A more intuitive way to evaluate the role of 'need' on regional tonsillectomy rate variation is displayed in table 21. In this table, the raw tonsillectomy rates over 10 regions can be seen, but also the tonsillectomy rates as a proportion of the population with recurring sore throat infections. And whilst there is great disparity between regions in the number of tonsillectomies undertaken, the proportion of tonsillectomies from a population of those with recurring tonsillitis in comparable.

Characteristic	Multivariable analysis	
	Adjusted IRR	P
Gender		
Male	1	<0.001
Female	1.39(1.24-1.55)	
Age category		
16-24 years	1	<0.001
25-44 years	0.16(0.14-0.17)	
Ethnic origin		
White British	1	<0.001
Indian	0.85(0.54-1.34)	
Black African	0.40(0.23-0.69)	
Black Caribbean	0.66(0.37-1.18)	
Black other	0.53(0.28-0.99)	
Bangladeshi	1.63(1.01-2.64)	
Pakistani	0.85(0.52-1.40)	
Other Asian	1.16(0.50-2.72)	
Chinese	0.57(0.21-1.52)	
Mixed	0.79(0.50-1.27)	
Other	0.57(0.39-0.85)	
Unknown	0.79(0.70-0.89)	
Social Deprivation		
Least deprived	1	0.67
Most deprived	1.06(0.92-1.22)	
Respiratory illness		
Absent	1	<0.001
Present	1.18(1.07-1.29)	
Obstructive sleep apnoea		
Absent	1	<0.001

Present	2.59(1.59-4.23)	
Obesity		
Not coded	1	<0.001
Obese	1.58(1.39-1.80)	
Alcohol		
Non-drinker	1	<0.001
Mild-Moderate drinker	1.26(1.14-1.40)	
Heavy drinker	0.61(0.47-0.78)	
Smoking		
Non-Smoker	1	<0.001
Ex-smoker	1.45(1.24-1.69)	
Smoker	1.30(1.20-1.42)	
Practice region		
North East	1	0.38
North West	0.88(0.57-1.38)	
Yorkshire & The Humber	0.97(0.59-1.59)	
East Midlands	1.23(0.75-2.02)	
West Midlands	0.91(0.59-1.40)	
East of England	0.92(0.59-1.45)	
South West	0.89(0.58-1.39)	
South Central	0.90(0.56-1.42)	
London	0.78(0.50-1.22)	
South East Coast	0.86(0.54-1.35)	

Table 21 Multi-variable models of tonsillectomy in adult.

*This table shows the risk factors associated with tonsillectomy, amongst **all adults 16-44 years old** in the CALIBER cohort, denominator=7.7million person-years. Multivariable models are based on the Poisson distribution. Multivariable models are multi-level and include random and fixed effect*

Once the above factors are accounted for there is <0.001 of the variance left unexplained at the level of region, and 0.23 of the variance still unexplained at the level of the general practice.

Practice region	Adults with recurring sore throat patients	Adult receiving tonsillectomy for recurring tonsillitis	Proportion
North East	128	63	49%
North West	3,541	1317	37%
Yorkshire & The Humber	956	382	40%
East Midlands	978	338	35%
West Midlands	3,130	886	28%
East of England	2,948	904	31%
South West	2,507	910	36%
South Central	2,019	673	33%
London	2,246	597	27%
South East Coast	1,970	710	36%
Mean	2042	678	35%

Table 22 Regional variations in tonsillectomy proportions for adults with recurring sore throat.

This table shows the proportion of adults who go on to receive a tonsillectomy having presented with a recurring sore throat, n= 22,465 patients with recurring sore throat.

Different measures of variation across health care settings						
	Children sore throat	Children rec sore throat	Children tonsillectomy	Adult sore throat	Adult rec sore throat	Adult tonsillectomy
% variation between highest and lowest region	33%	86%	290%	24%	50%	21%
Greatest univariate odds ratio	1.39(1.31-1.47)	1.86 (1.50-2.31	2.13(1.42-3.18)	1.29(1.24-1.34)	1.48(1.24-1.77)	1.21(0.93-1.58)
Multivariate odds ratio for above comparison	1.37(1.27-1.47)	1.65(1.31-2.08)	1.92(1.30-2.84)	1.23(0.92-1.64)	1.44(0.98-2.10)	1.23(0.75-2.02)
Unadjusted SCV	4.2	4.8	8.8	3.4	4.2	5.3
Adjusted SCV	1.2	1.8	2.3	1.7	2.1	2.9

Table 23 Different metrics of variation as measured across health care settings

This table shows the different metrics used to measure variation for each of the subject types. The variation is greatly reduced when using adjusted SCV.

Discussion

Summary

I have shown that there is substantial unadjusted variation between regions e.g. in children there was up to 35% variation between regions in GP sore throat consultation (unadjusted SCV=4.2), 90% GP recurring sore throat consultation (unadjusted SCV= 4.8) and 290% tonsillectomy rates (unadjusted SCV=8.8). A wide range of risk factors for sore throat, recurring sore throat and tonsillectomy were identified. (see Table 23 Different metrics of variation as measured across health care settings). Once these patient variables were accounted for, geographical variations in sore throat, recurring sore throat and tonsillectomy rates reduced and SCV measurements based on these models suggested only minimal regional variation after adjusting for patient characteristics and regional patterns of recurring sore throat, with an adjusted SCV of 2.3 for tonsillectomy in children and 2.9 for tonsillectomy in adults.

Strengths

This is the first study to investigate the effect of patient level characteristics on regional variation of tonsillectomy rates. By using a dataset that linked primary and secondary care encounters, I was, for the first time, able to relate regional tonsillectomy rate variations to population predictors of regional sore throat disease burden in primary care. Firstly, I could reliably investigate geographical variations across the patient care pathway, from the initial sore throat consultation in primary care through to tonsillectomy in hospital. Secondly, I had a dataset that had more complete information on the tonsillectomy (secondary care database) linked to a dataset that had more complete information about patient socio-demographics and co-morbidities (primary care database), allowing us the opportunity to understand the role of patient characteristics on regional tonsillectomy rate. Additionally, I used a multi-disciplinary team to define the codes that would phenotype our outcomes of interest in an iterative and transparent manner. Finally, the use of electronic health care records, which were collected as part of routine medical care, allowed for long follow up (mean follow up 6.8years) which would be expensive if the data were collected specifically for a research study, and data collection without risk of changing participant behaviour that

would occur in a study setting. Finally, this is the first study that describes regional adult tonsillectomy rate variation, showing that the level of variation is substantially lower than in children (see Figure 9 Sore throat in the community to tonsillectomy in secondary care for children 5-15 years and Figure 10 Sore throat in the community to tonsillectomy in secondary care for adults 16-24).

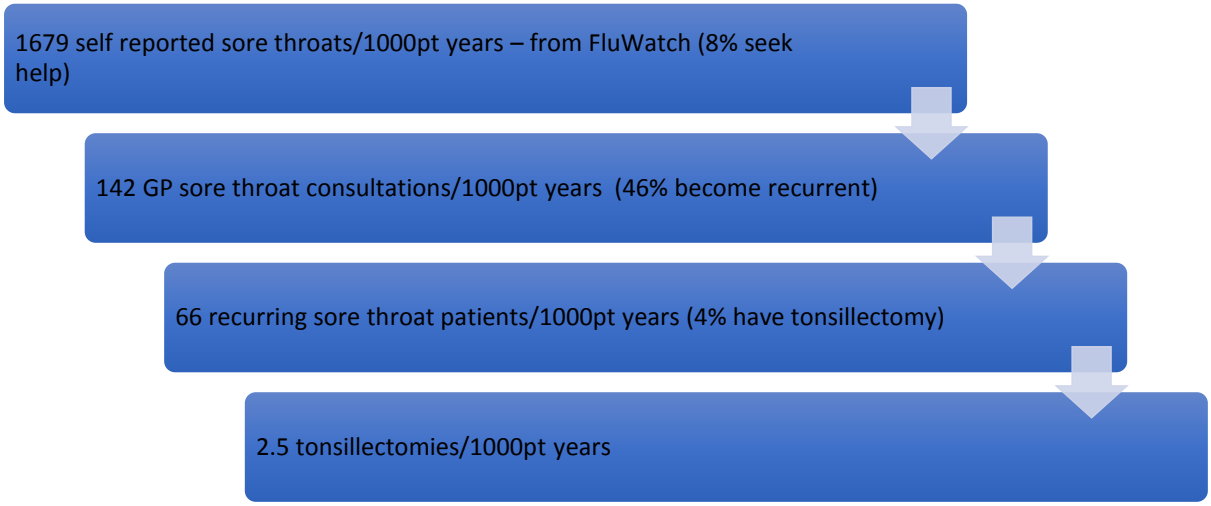


Figure 9 Sore throat in the community to tonsillectomy in secondary care for children 5-15 years

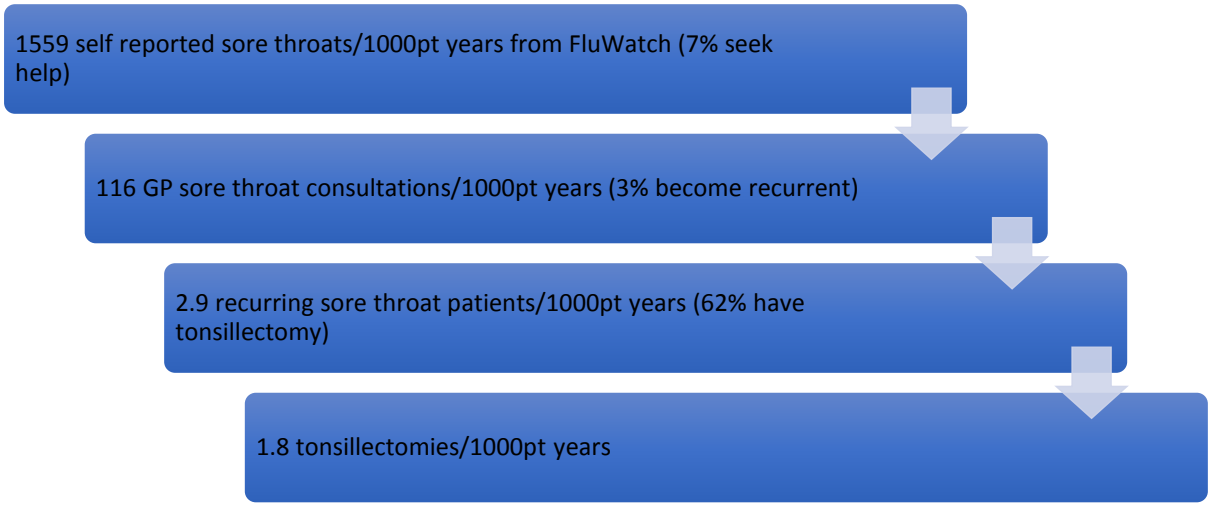


Figure 10 Sore throat in the community to tonsillectomy in secondary care for adults 16-24

Limitations

Whilst the use of electronic health care records had the advantages described above, it also had several disadvantages. Firstly, the database was not created with our research question in mind and so there were important pieces of information that I could not use since they were infrequently coded by GPs (e.g. the presence of a fever during a sore throat, severity of symptoms, referral to ENT, etc.). Additionally, whilst I could identify whether patients were part of the same GP and Strategic Health Authority, I did not have access to data that would allow us to identify the patient's Primary Care Trust/Clinical Commissioning Group (PCT/CCG) or even Hospital Trust, which is where local policy is enacted. Therefore, it would not be possible to say anything about inter-hospital variation or even comparisons between localities that have differing local policies on tonsillectomy. Finally, our dataset did not include Accident and Emergency visits for sore throat and so could not measure all health service activity related to sore throat. However, it must be noted from our FluWatch study that only a small proportion of patients seeking help for their sore throat attend A&E (3%), and so I did not expect the loss of this data to affect our overall conclusions.

Findings in the context of published literature

In response to a local study of regional variation in Oxford (19) Suleman (19) (2010) undertook an analysis of regional variation in paediatric tonsillectomies using HES and private health insurer (BUPA) data from 2000-2005 and reported a 7-fold difference between Primary Care Trusts (now called Clinical Commissioning Groups). The King's Fund published a report on regional surgical rate variation, using 2009-10 HES data they also reported a 7 fold difference in paediatric tonsillectomy rates between PCTSs (20). The Rightcare(175) programme have since done the same and reported a was a 5.2-fold difference between the PCT with the highest and lowest rates, based on 2011 HES data. Our data analysis showed that there was 3-fold difference in paediatric tonsillectomy rates between the regional levels I analysed (strategic health authority level) with the lowest and highest rates.

Some of the difference between the literature and the results reported here may relate to the different geographical units assessed. Due to pseudononymisation and patient re-identification policy around linked primary and secondary care datasets I did not have access to data that would identify patient's primary care trust or hospital. Studies described in the previous paragraph did not have access to individual patient data but only aggregate population data and so were able to discuss variations at local health authority level. Therefore, it is difficult to make direct comparisons between our study and the others described above, since they describe variations at different size geographical units. It would be expected that variation reduces when larger geographical units are analysed (e.g. strategic health authority compared to primary care trust) since the rates of outlier primary care trusts are more heavily diluted through regression to the mean. This is borne out in my results which show greater variations for GPs than regions. Our analysis showed that there were still large differences in tonsillectomy rates, even with a smaller dataset and analysis of larger geographical units. Most importantly my results demonstrate that whilst crude metrics of variation can provide measures of maximum variation, they may be misleading. In order to formally examine such apparent variation and it is essential to use SVC thresholds for categorisation.

No previous study has described regional tonsillectomy rate variation in adults. It is unclear whether this is related to publication bias (with a greater likelihood to report high levels of variation than low levels of variation) or a perception that adult tonsillectomies are infrequent. Although adult tonsillectomy rates are substantially lower in adults than in children it must be noted that nearly half of all tonsillectomies done in our cohort were on adults.

The King's Fund reported a systematic component of variation of 8.4 in relation to paediatric tonsillectomy rate variation between local health authorities (20), which is comparable to unadjusted SCV of 8.8 that I report here. The systematic component of variation is a more robust measure of small-area variation, as it focuses on inter-regional disparities, after accounting for expected variation due to population characteristics. When I recalculated the SCV using a more robust statistical model (which included patient level predictors of

tonsillectomy and clustering of patients to general practices) the SCV dropped to 2.25. Whilst the authors of the King's Fund publication do not describe the underlying model of their SCV calculation, it is unlikely they had access to accurate patient level characteristics since they used an unlinked HES database in isolation.

My FluWatch study (chapter 3), along with RightCare' s Atlas of Practice Variation (59,176) adds further support to the hypothesis that regional tonsillectomy rate variation is substantially affected by regional 'need'. My study of CALIBER data showed East Midlands had the highest rates of tonsillectomy, whilst North East England had the lowest rates. The variation between these two regions was most apparent in children (0-15 years old) and had an almost 300% difference in tonsillectomy rates between these regions, compared to adults (16-44years old) who had 20% difference between these regions. Regional patterns of recurring sore throat (3 primary care sore throat consultations in less than 12 months) showed residents of the East Midlands were more likely to have recurring sore throat compared to those in the North East of England (children 90% more likely and adults 50% more likely). These observations are corroborated by my FluWatch study (chapter 3) that showed regional patterns of community sore throat residents of East of England and East Midlands were more likely to self-report a sore throat compared to those in North England (children 3% and adults 14% more likely). Overall, regional patterns of sore throat in the community, and incidence of recurring sore throat in primary care all seem to vary in patterns that are similar to regional tonsillectomy rates. This holds across 3 separate databases: FluWatch, CPRD and HES. This suggests that a considerable component of regional variation in tonsillectomy rates may be present before the patient is seen in an ENT outpatients setting.

Future work is also planned to investigate regional variation in other surgical procedures such as grommet insertions for otitis media with effusion as mechanisms that drive variations in other procedures may be different.

Implications for policy

Current strategies to reduce regional tonsillectomy rate variation may have failed to reduce disparities as they failed to realise a significant proportion of this variation relates to regional

variations in recurring sore throat, and potentially community incidence of sore throat. Regional tonsillectomy rate variations may be better described as a symptom of a deeper issue that allows for regional disparities in healthcare. Future policies should promote joined up strategies to reduce variation in care across the entire care pathway, from sore throat management in primary care through to surgical management with tonsillectomy in secondary care.

A key component of the current Government policy to reduce regional surgical rate variations is to quantify and publicise variations through RightCare's Atlases of Practice Variation (59,177,178). This allows local healthcare authorities to judge over- or underuse of treatments and make appropriate changes through prioritisation (rationing) strategies. However, these Atlases are based on unadjusted figures, taken at from aggregate HES data. As this study has shown, unadjusted regional rates of tonsillectomy, may distort true variations in tonsillectomy rates and result in ineffective use of commissioning guidance, which at best may have no effect on regional tonsillectomy rate variation, but at worse may prevent patients with recurring tonsillitis failing to receive appropriate treatment.

Additionally, whilst the RightCare Atlas of Practice variation described regional paediatric tonsillectomy rate variation my freedom of information request showed local commissioning guidance applied to adult tonsillectomy also. This is the first study that has reported on regional adult tonsillectomy rate variation and has found there is very little variation in this group. Therefore, local policy aimed at this population may be unnecessary, and may, in fact, be harmful.

CHAPTER FIVE: Can aspects of the decision-making process contribute to regional variation in surgical rate?

Chapter synopsis

In this chapter I have introduced possible alternative sources of variation in adult tonsillectomy by considering the variables that can be measured within the medical consultation. The chapter provides an overview of the main conceptual frameworks of shared decision making –the most relevant description of the medical consultation between doctor and patient (i.e. decision support, information exchange, patient centredness, patient empowerment, decisional uncertainty, role preference, patient treatment preferences). It goes on to review the extent to which these key concepts can be measured, and explains the reasons for measuring decisional uncertainty and treatment preference in an observational study of decision making for adults with recurring tonsillitis.

Introduction

Chapters 3 and 4 have shown that the rates of sore throat in the community and in primary care, along with tonsillectomy rates in secondary care change with factors that are known to vary between regions (specifically age, sex, ethnicity, presence of chronic medical diseases, number of people in the household, smoking status and population density). Controlling for these population factors reduced regional tonsillectomy, recurring sore throat and community sore throat variations, with the SCV reducing from 8.8 to 2.3 in children's tonsillectomy rates and 5.3 to 2.9 in adult's tonsillectomy rates.

The greatest evidence for regional variation comes from the concept of the "surgical signature". The surgical signature refers to regional variations in surgical treatment, corresponding to the differences in surgeon preference for certain treatments. Yet, there is very little evidence for why shared decision making in one region leads to patients choosing tonsillectomy and others not. Some have suggested (39,40,60,179) that other factors, at the level of the individual patient, the individual doctor (physician/surgeon) or the interaction between the two during the medical consultation (collectively defined as the medical decision-making process) can also contribute to variation in tonsillectomy rates. Ideally, a study of decision making in relation to regional tonsillectomy rate variation would have examined variation at the level of GPs in primary care and hospital (surgical team). We do not know which part of the shared decision-making process could be related to decisions becoming regionally aligned. A better understanding of the conceptual frameworks leading to medical decision-making is required in order to target interventions aimed at reducing variation. Measuring decision making in primary care over several consultations would have been difficult conceptually, as well as practically. Before I embarked on a study to understand if any parts of the medical consultation for recurring tonsillitis had any role in regional tonsillectomy rate variation, I set out to understand which aspects of the medical decision-making process are most likely influential and which aspects can be practically measured. As a first step this was most feasible in the consultation between surgeon and patient, which led to a measureable decision outcome (tonsillectomy or watchful waiting), between only 2 agents.

Shared Decision-Making (SDM): Definition and key conceptual frameworks

Since the time of Hippocrates, the medical consultation has followed a doctor-centred "paternalistic" model, defined as the doctor giving the patient selected information and encouraging the patient to consent to what the doctor considers best(180). However over the last half-century, due to a shift in focus of medicine from acute to chronic diseases (181), increase in the number of treatment options available(63) (182), and changes in society that have informed (183) and empowered individuals (184), there has been a move away from this model of decision-making, towards a more patient-centred approach, defined as shared decision-making (SDM).

Charles and Gafni (185) originally described patient participation as shared decision-making (SDM), and defined it as a process that has 4 features:

1. At least two participants: the doctor and patient;

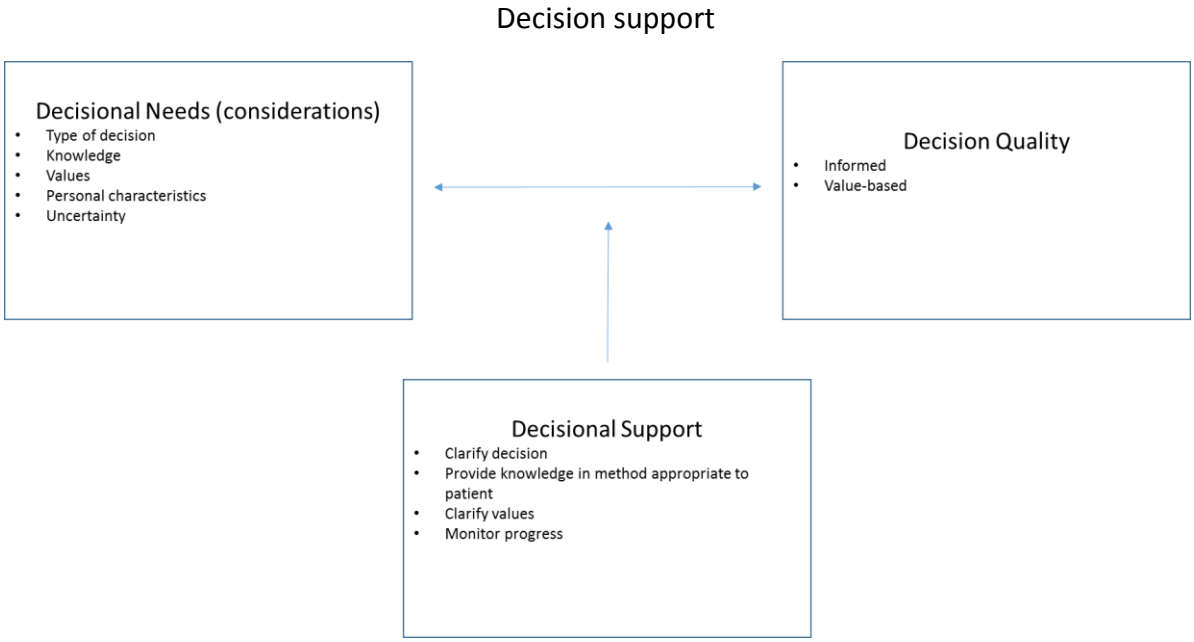
2. Information is shared;

3. Both parties (doctors and patients) take steps to participate in the process of treatment decision-making;

4. A treatment decision is made and both parties agree to the decision.

SDM is now extensively taught to medical practitioners training and practicing in both primary (186), (187) and secondary care (188,189). In parallel, there has been a national campaign, aimed at the public, that encourages a change in the method of medical decision-making, away from the traditional "paternalistic" model towards to a more shared approach(190). Whilst the popularity of SDM, has grown so has its definition. A recent review (191) found 161 different definitions for the process, suggesting that the term SDM is an "umbrella" term incorporating many different, specific concepts related to patient-centred decision making. Conceptual frameworks (concepts or constructs) are definitions of abstract ideas, based on a structure of relationships between key-variables(192). There is a difference in technical language depending on the background and discipline of authors. For this reason, the term concept and construct will be used interchangeably throughout this chapter.

Since concepts around the process of SDM have been developed independently of each other, it is unclear how these concepts relate to each other within the process of SDM. Further, there is little clarity about the names of these specific concepts since concepts are frequently referred to using different names. To understand more fully the range of concepts that are discussed in relation to SDM, and to what extent they may relate to one another to create the process of shared decision-making I undertook a review of the relevant conceptual frameworks. Through brief review of the relevant literature (search terms in PubMed and PsychINFO: “shared decision making”, or “SDM”, or patient-centred decision-making” or “patient-centred decision-making” and “conceptual framework” or “concept*” or “construct”) I identified seven main concepts in relation to the processes of SDM: decisional support, patient centredness, information exchange, empowerment, decision uncertainty, role and treatment preference. Each of these contains several separate elements and there is considerable overlap between them. Below, I discuss the main conceptual frameworks that have been used to describe these ideas.



The Ottawa Decision Support Framework (

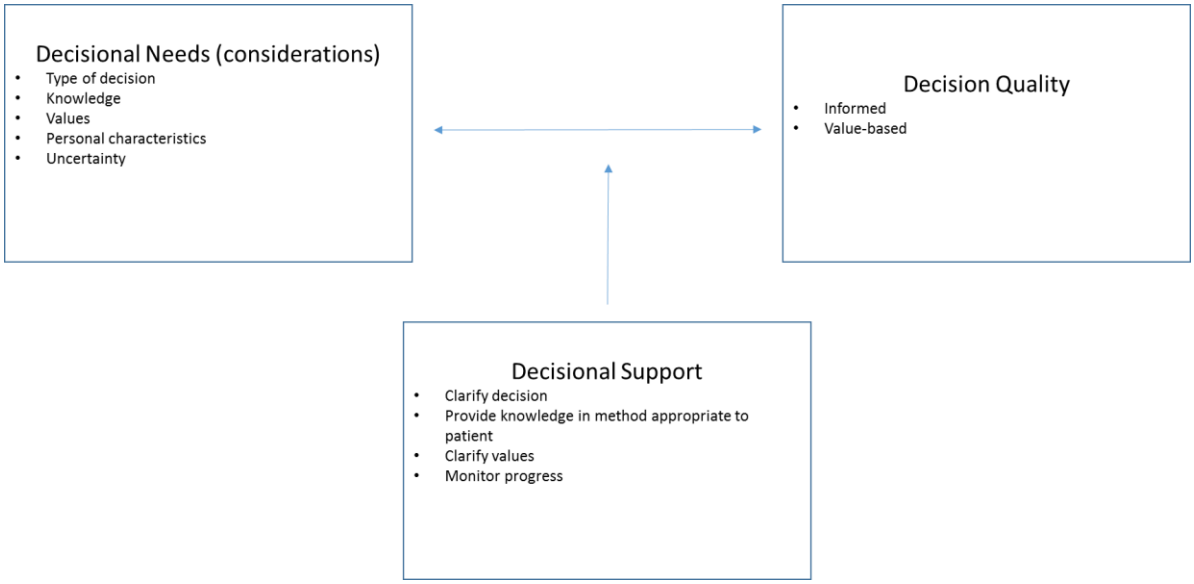


Figure 11) provides a structure to support patients through a medical decision, based on theories from general psychology (193) social psychology (194), decision analysis(195), decisional conflict(196), social support (197) and economic concepts regarding expectations and values(198). The framework describes supporting a patient through their medical

decision by clarifying their need, providing facts and probabilities, clarifying their values, guiding them in deliberation and then finally helping facilitate their decision.

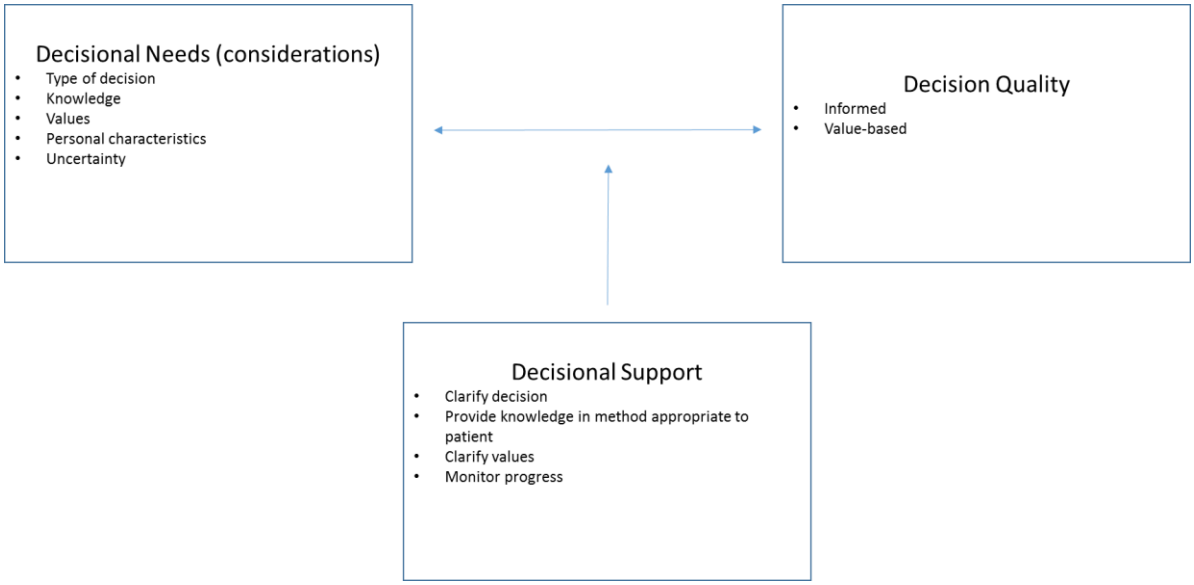


Figure 11 Decision Support, my adaptation based on Ottawa decision support framework(199)

The decision support framework gives a structure to support patients through making medical decisions above shows how the authors describe decision support from the patients' decisional needs to a high-quality decision

Information exchange

Frederickson's (200) framework (Figure 12) provides an in depth analysis of information exchange, building on Tuckett's (201) five general statements of information sharing for the general physician recognising that: (1) attending to the provision of information, reassurance and understanding are intrinsic and important parts of therapy; (2) the patient's co-operation in carrying out advice cannot be taken for granted and information may be necessary for persuasion; (3) the outcome of medical treatment is multidimensional and subjective matter, thus the patient's view is relevant; (4) individuals selectively seek help with symptoms that they experience and seek knowledge to deal with them; (5) patients are in a sense consumers of the medical service and require information and autonomy in decision making.

Frederickson L (202) realised that the concept of information should extend beyond the mere facts of the disease and should include insight into the attitudes, feelings, fears, desires,

expectations and anticipations of the patient and should be pervasive through the whole consultation.

Frederickson described three key dimensions that exist for both patient and doctor separately: Input; Information exchange; and Outcomes. “Input” relates to the relative knowledge, orientation, perspectives, social norms and belief systems that exist within and around each agent, prior to the exchange. As part of the “exchange” process Frederickson (202) described the patient and doctor role introductions (203). They allowed for a fluid model of information flow to mimic the unpredictable flow of human interaction. In addition to verbal communication they also felt that the concept of information exchange should not be limited to verbal exchange. Intrinsic to the concept of information exchange was the notion of information processing by the participants, and this subsumed the perception, selection, and understanding of informative items. Significant processing was necessary before any information available in a consultation could be utilised by those who received it (204). The process depicted in the model readily incorporated the patient-centred ideology of Levenstein et al. (205) and Middleton (206) with the information exchange involving exploration of doctor and patient agendas.

Finally, “output” describes the full disclosure of treatments being considered and commitment to options offered by one another. Figure 12 below illustrates this process of information exchange. This construct is similar to the idea of knowledge transfer described by Lavis et al as producer push, user pull, information exchange process(207).

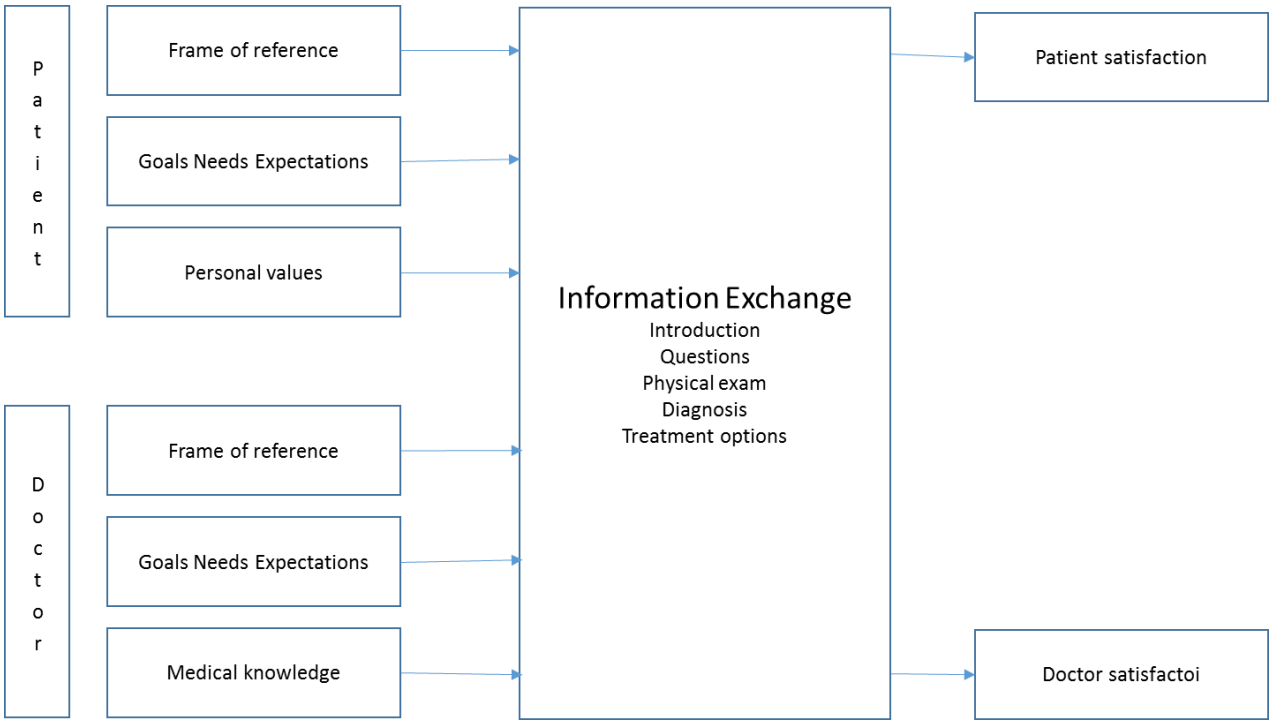


Figure 12 Information exchange – diagram based on Fredericksson (200)

This framework, based on Fredericksson, illustrates the process of information exchange between a doctor and a patient.

Patient centredness

Mead et al (208) derived a formal framework of ”patient-centredness” based on a review of conceptual and empirical literature regarding the doctor-patient relationship. The authors described five key dimensions of patient centredness (Figure 13 Patient centredness) as the biopsychosocial perspective, patient-as-person; sharing power and responsibility; therapeutic alliance; and doctor-as-person perspectives. The biopsychosocial perspective, derived from work from (209), broadens the definition of illness from just the biomedical to include psychological and social aspects in a hierarchical manner. This model can best be understood with the example of a man who died suddenly of a heart attack at work. The biomedical model would suggest that he had a heart attack because a blockage in the vessels

taking nutrients to the heart. The biopsychological model would describe this situation as a stoic gentleman that has been having chest pain related to narrowing of his heart vessels for months and has refused to seek help, culminating in a complete blockage and a heart attack. The biopsychosocial model would describe this situation as stoic gentleman that has been having chest pain related to narrowing of his heart vessels for months and has refused to seek help, culminating in a complete blockage and a heart attack that could not be treated urgently enough due to lack of defibrillators at work. While the biopsychological model included a broader range of concepts than the biomedical model, Mead et al (208) felt that even the biopsychosocial model did not sufficiently describe the situation as it did not account for the patients' perspective of illness. The patient-as-person, therefore, adds a biographical component, created to instil a sense of illness within the patient's life setting. For example, in the above case it would provide information on how the patient viewed his symptoms, illness and death. Mead added "sharing power and responsibility" to the conceptual framework of patient centredness based on Byrne and Long's (210), analysis of audiotaped consultations. Byrne and Long described sharing power and responsibility as any process that results in recognition of patients' needs and preferences in their treatment. Mead added "therapeutic alliance" based on new developments from Roth & Fonagy (211) to emphasise certain aspects of the patient-doctor relationship : (a) the patient's perception of the relevance and potency of interventions offered, (b) agreement over the goals of treatment, and (c) cognitive and affective components, such as the personal bond between doctor and patient and perception of the doctor as caring, sensitive and sympathetic. The final dimension added to this construct was doctor as person, to acknowledge doctor subjectivity was an inherent component of the patient-doctor interaction (212). Mead suggests that five key definitions of patient centredness relate to each other through information exchange, but does not qualify the basis or types of interactions involved. Patient involvement is often used interchangeably with patient centredness. Whilst patient involvement was first described in 1974 and formalised through the creation of community health councils, this term has also been defined from a health policy perspective to help align health services to better match

patients' needs (213) . Currently it has since become synonymous with involving patients in all their health care decisions (214).

Figure 13 Patient centredness below describes the potential contributors to patient centredness, which are divided into patient, doctor and consultation factors, all of which can be influenced by social and professional norms.

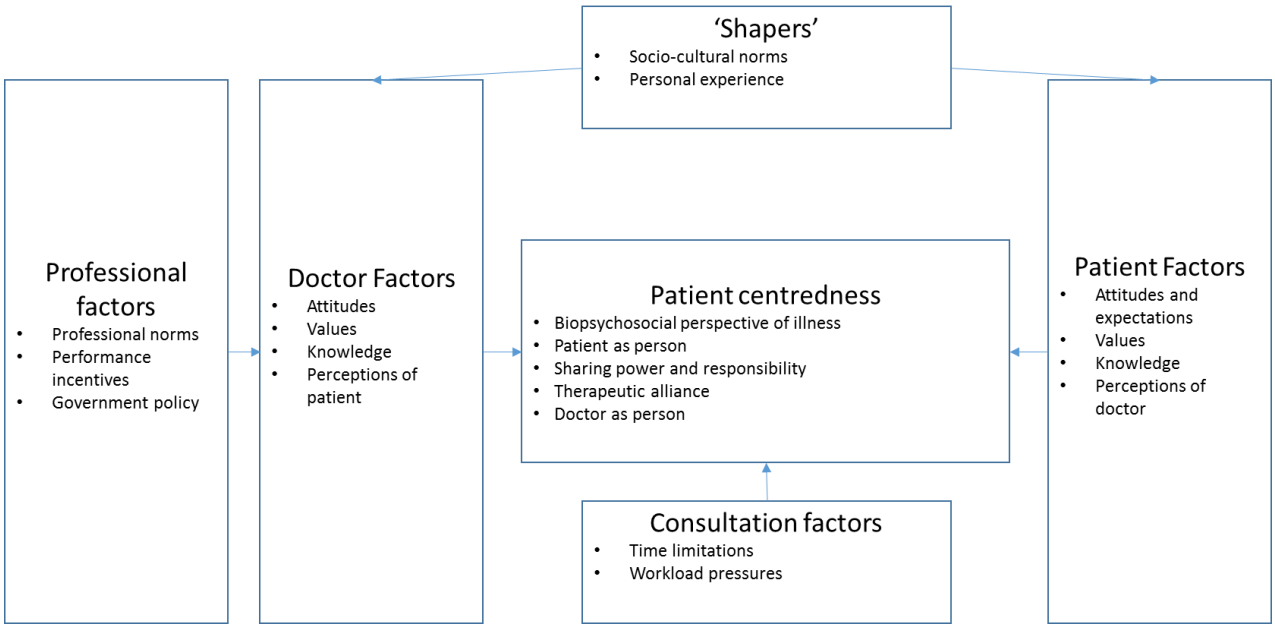


Figure 13 Patient centredness

This figure shows my adaptation of the components which contribute to patient centredness, based on theories by Mead {Mead:2000cb}

Patient empowerment

Opie (215) reviewed previous conceptual frameworks of patient empowerment over material circumstance (216), empowerment over social circumstance (217) and the method by which professionals can empower patients (218). The authors then undertook qualitative interviews to investigate how multidisciplinary teams could best be deployed to empower patients and their carers. Menon (219) built on these frameworks to describe empowerment as perceived control, perceived competence and goal internalization.

Goal internalization of health ideals requires that patients see good health as their personal goal. Perceived competence relates to the patient's management of various activities associated with health and health care. This requires patients to a) have the capability and knowledge required to maintain a healthy lifestyle, b) have the ability to manage minor ailments that do not require specialised medical assistance, and c) know when, where, and how to seek specialised medical assistance. Perceived control relates to controlling factors associated with maintaining good health. This frequently requires access, high quality of services and interaction with health care providers.

Decisional uncertainty/conflict

The concept of decisional uncertainty/conflict grew out of studies examining how people coped with stresses in World War II and culminated in the "conflict-theory model of decision making"- see Figure 14- (220). The authors felt that uncertainty/conflict was the result of an opportunity or need to change the status quo, insufficient knowledge about benefits and risks of continuing in status quo or undertaking the alternative, and insufficient time to evaluate and undertake a change.

The five coping patterns identified when facing stresses of uncertainty/conflict were un-conflicted adherence, un-conflicted change, defensive avoidance, hypervigilance and vigilance. Studies investigating physiological concomitants found that the greater the uncertainty/conflict the more marked the changes in heart rate and galvanic skin-response (221,222)

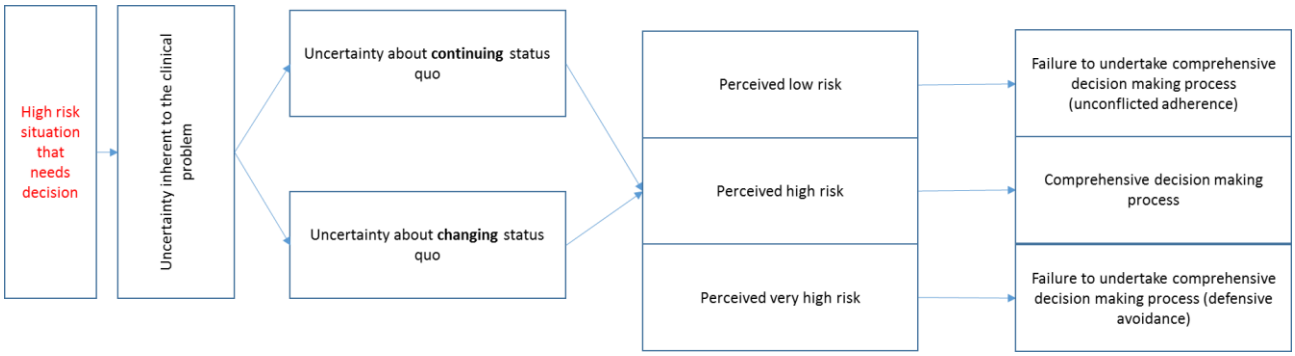


Figure 14 Conflict-decision theory Adapted from Janis and Man (223)

This model shows the process used to reduce uncertainty and conflict when making a decision.

The model of uncertainty/conflict (Figure 14) describes the process required to appropriately reduce the uncertainty/conflict in a decision, i.e. 1. Information gathering about potential gains and losses from change in status quo; 2. Information gathering about gains and losses from alternatives; 3. Having sufficient time to undertake evaluation of above. Whilst the original model was created to understand decision-making in uncertainty, it was not until O'Connor {OConnor:1998hh} used the model to understand erratic health choices for breast cancer and vaccination that it was used to describe and evaluate how shared clinical decisions were undertaken. Although the work of Janis and Mann (220) was seminal, modern neurobiological research on conflict and decision-making also considers heuristic processes rather than deliberate-rational processes alone(224).

Role preference

The idea of role preference originates in the work of Degner(225), who undertook a qualitative study designed to investigate "What happens when treatment decisions are made for patients with life-threatening illnesses". Degner used role and control preferences interchangeably throughout their article. The authors reported that whilst several factors influenced the way treatment decisions were made, the central factor appeared to be "control over the design of the treatment." There were four main patterns of decision making: provider-controlled, patient-controlled, family-controlled and jointly controlled (225). This led the authors to a second question: "Do patients actually have preferences about the degree of control they actually want to exercise in treatment decision-making?" Analysis (using grounded theory) of consultation transcripts from first visits made by patients to two different cancer clinics revealed that patients fell into 3 categories: delegators (those that delegated the decision to the doctor), deliberators (those who wanted time to go away and think about the information before coming back and making the decision with their doctor) and decision makers (those that wanted to make their own treatment choices) – see Figure 15. In an effort to decide whether patients could differentiate their preferred roles in treatment decisions Degner asked 60 patients from oncology clinics to sort through 4 cards that represented potential roles (226)– see Figure 15. The authors reported that most patients chose to share their decisions, and therefore they added a fifth option to help

differentiate amongst those who chose to share - Figure 15. Degner went further to show that preference for role in decision making was separate from preference for information regarding treatment options (225). Sutherland, Llewellyn-Thomas, Lockwood, Richler and Till (227) have since commented that need for information is related to the need to enhance psychological autonomy and is not necessarily related to a desire to assume responsibility for the treatment decision. Finally, Degner (228,229) used a survey studies of breast cancer patients to assess whether patients were able to differentiate between preferred decision roles and actual treatment roles. The studies showed that patients were able to differentiate between the two, and that patients who had been actively involved had a higher quality of life. Degner suggested that role preference was dependent on patients' beliefs in the efficacy and benefits of self-care and should be communicated between patient and doctor. The term role preference is interchangeably used with control preferences throughout the literature.

Figure 15 Role preference construct Adapted from Degner et al (230)

This figure shows the different types of roles which patients may choose to take in a consultation.

Patient treatment preference

A key additional characteristic that patients bring to the clinical encounter is their preference for treatment. The psychological process that converts information gathered into a preference has historically been modelled using economic or psychological disciplines. Economic theory bases preference on utility, which is a measure of the satisfaction gained from the consumption of a good or service (231) – or in health care, the treatment decision. In psychology, preference is measured in the concept of attitude, which is defined as a disposition of a person to respond favourably or unfavourably to an object, person, institution

or event (232). Both of these models are based on the paradigm of a rational, individual decision maker and are concerned with two key processes: (1) An evaluation of an intervention in terms of its desirability or attractiveness and (2) A choice between alternative interventions based on that evaluation (233). Asking patients about their preferred treatment should initiate both processes. However, as complexity of information increases so does the probability that the respondent departs from the rational decision-making and employs simplifying heuristics to form a preference, and therefore, omits the evaluation phase (234). Preferences measured this way are not always stable with time (235,236).

Review and inter-relatedness of conceptual frameworks

All constructs related to the process of SDM and all frameworks described above aimed to understand patient and doctor collaboration. None of the constructs described the outcomes of SDM, such as satisfaction, or knowledge retained by the patient. Patient-centredness, empowerment, and decision support related more to the doctor's role in the process. Information exchange included both patient and doctor's roles. Role preference, decision uncertainty and treatment preference were more patient-orientated constructs (i.e. they related to patient's role in SDM). Critical appraisal of each of the constructs above), suggested that there may be some conceptual overlap between constructs, although different terms were used to describe them. For example, conceptual definitions of decision support, decisional uncertainty and patient centredness all include information delivery, which is an integral part of the concept of information exchange.

To investigate the interconnectedness of SDM constructs I examined each construct definition again to see if it could fit into the broader definition of any of the other constructs. I numbered key definitions of each construct and compared them to numbered key definitions from the remaining constructs. For example, "*Patient involvement in decisions"* was a key definition in the conceptual framework of Patient Involvement. I felt that this definition was related or similar to "*value elicitation*" in Decision Support, because if patients' values are being elicited they are being involved in their decision. I also felt it was similar to "*Discuss effect of problem on patient's life*" in Information Exchange, because by discussing the effect of problem on patient's life the patient is becoming more involved in their

treatment decision. By repeating this process across all key definitions of the seven conceptual frameworks I could hypothesize how the conceptual frameworks were related to each other. The output of this process is summarised in Appendix V – Table of Interrelatedness of Concepts – and described in an example below the table. This quantifiable process allowed me to create a representation of how the conceptual frameworks related to each other.

I interpreted this table (Appendix V – Table of Interrelatedness of Concepts) as showing three constructs that were connected to nearly all other conceptual frameworks (patient involvement, decision support and information exchange). In contrast, there were four constructs that had little commonality between them (i.e. treatment preference, role preference, decisional uncertainty and empowerment). For example, neither empowerment nor role preference had any association with decisional uncertainty or treatment preference.

I used an iterative discussion to develop a diagram of how all the constructs related to each other within the overall process of SDM. Part of this iterative discussion related to whether a construct was a smaller integral or a larger overarching construct. Through this process, I concluded that there were two large overarching constructs of SDM: patient involvement and decision support, each including several of five smaller constituent constructs: information exchange, role preference, patient empowerment, decisional uncertainty and treatment preference. Information exchange was the most inter-related of all constructs, having defining features in common with all six other constructs. Therefore, information exchange was described as a central construct, one that allowed all the other constructs to relate each another. Figure 16 presents these inter-relationships diagrammatically.

Summary

All constructs discussed above have some foundation in theory that has been formalised into definitions using observational studies. Some definitions are broad and include many aspects of SDM (information exchange, patient centredness and decision support), whilst others are very specific to a small but integral part of SDM (role preference, treatment preference, decisional uncertainty and patient empowerment). It can be argued that larger, indistinct

constructs are more difficult to measure using questionnaire instruments. Many of these constructs were defined before SDM was described and, therefore, there is considerable overlap between constructs, in the description of SDM.

Given the inter-related nature of these constructs, care must be taken when choosing a construct to measure. In a study that is designed to investigate the role of the SDM on treatment chosen for recurring tonsillitis, measuring a small, discreetly defined construct (e.g. role preference) may provide results that are more readily interpretable and translatable into clinical practice (e.g. improve opportunities for patients to understand and state their preferred role in the decision). On the other hand, it may miss a large association between SDM and treatment chosen. Conversely, measuring a larger, indistinct construct (e.g. information exchange), may identify an association between the SDM and treatment chosen, but it may render interpretations difficult to translate into clinical practice (e.g. should one increase the amount of information, the type of information or the format the information is deployed in). Therefore, it is essential that I next considered how measureable these constructs were in the context of study that assesses shared decision making for adults with recurring tonsillitis. In the following section, I have examined existing questionnaire instruments that claim to measure constructs of the SDM process. The following chapter will focus on critical evaluation of methods for measuring treatment preferences and the development of an instrument that can undertake this for adult tonsillectomy.

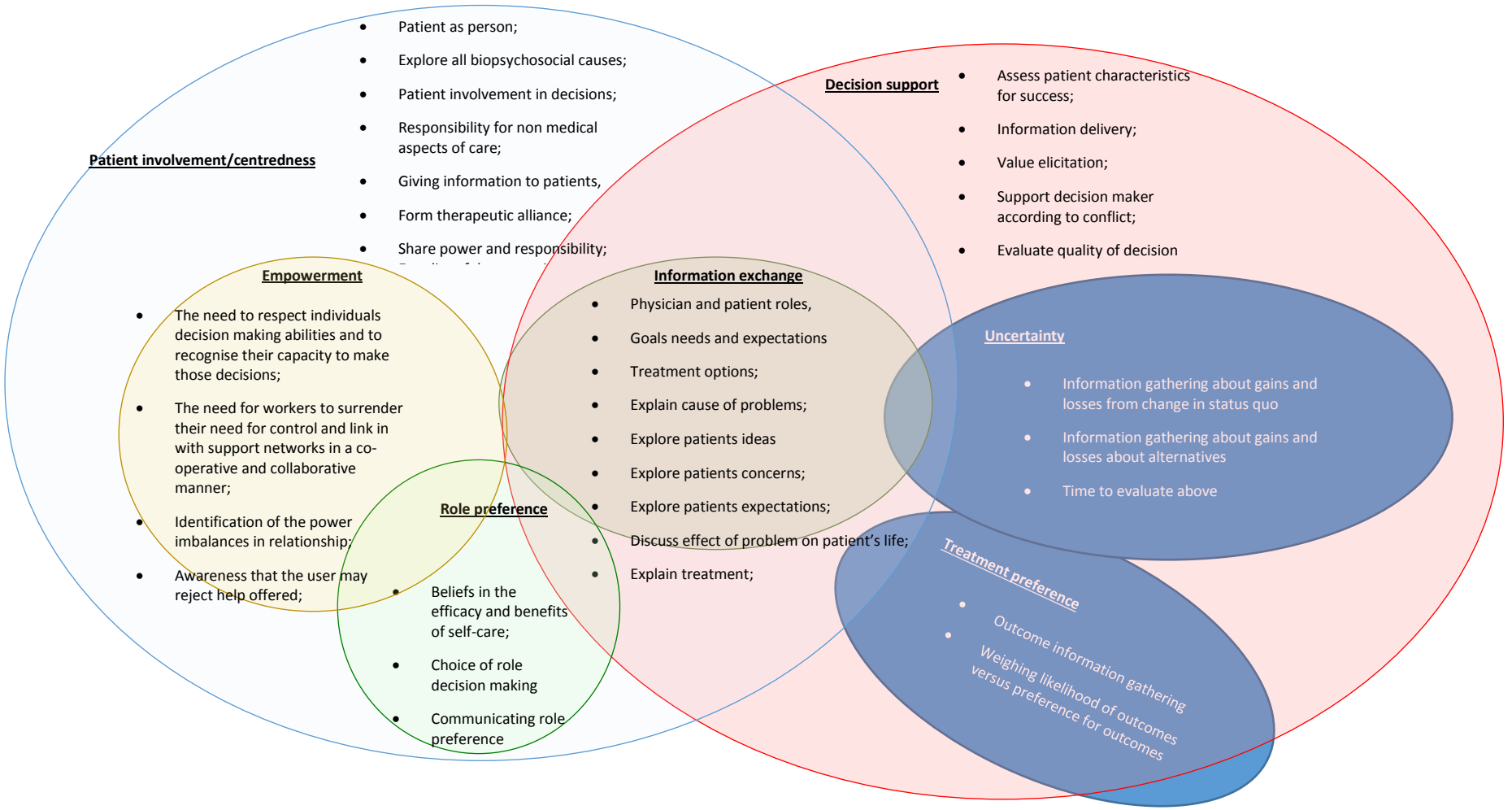


Figure 16 Overlap of concepts of SDM.

Questionnaire instruments that measure constructs of SDM

Having identified the key concepts that are relevant within the decision-making process (decision support, patient involvement, information exchange, role preference, empowerment, uncertainty and treatment preference); for those to be potentially useful in investigating variation they also need to be measurable. To address this question I identified the existing questionnaire instruments that claimed to measure one of the above conceptual frameworks (or closely related) and critically appraised the psychometric properties of each using criteria with adapted from several well recognized sources (237-240).

A review of the literature identified instruments that measure constructs of SDM (PubMed and PsychINFO were searched using the search terms "shared decision making", or "SDM" or "patient involvement" or "patient centred" or "patient centred" or "decision support" or "role preference" or "treatment preference" or "decisional uncertainty" or "decision uncertainty" or "decisional conflict" or "decision conflict" or "empowerment" or "information exchange" or "knowledge transfer" cross referenced with "instrument", "tool", "survey", "questionnaire" and "measure").

Original search for instruments that measure constructs of SDM revealed questionnaires that could be divided into groups based on: 1) Whether they measured the process or outcome of SDM; and 2) The perspective they measured this from (patient, doctor, independent observer). I focused on measures of decision making process itself (e.g. how much information exchange occurred during SDM), rather than satisfaction from the decision-making (e.g. how much satisfaction resulted from the information exchange that occurred during SDM). This is because studies have shown that patient decision satisfaction can be easily conflated with rapport with doctor or satisfaction from the treatment itself(241) ((242-246) (247) (248)).

I also restricted instruments by the perspective they used to measure their construct (i.e. doctor-reported, patient-reported or independent observer reported). Studies have shown that when a construct is evaluated from different perspectives the results are not interchangeable or even comparable(249) (250) (251) (245,252) (253) (254) (255).

Changing the perspective from which the measurement is made may slightly change the underlying construct being measured. Independent- observer based instruments miss information that is non-audible or is non- observable. Additionally, they do not take account of the *perceptions* of those who are involved in the process of understanding the nature of decisions, negotiating their role in the decision process, and acting to deliberate and decide. Additionally, nearly all constructs described above represented conceptualisations of SDM that were best reported from a patient's perspective (e.g. how involved was the patient, had the patient described their preferred role, had the patient appropriately chosen their preferred treatment, etc.), and therefore patient reported instruments were more likely to align with the underlying construct. For all the above reasons, I focused on patient-reported instruments only.

Scientific criteria are available for evaluating the psychometric rigour of rating scales (i.e. questionnaires). For example, gold-standard review criteria have been published to evaluate the scientific and practical aspects of health outcome measures(240). Guidelines for the development, testing and dissemination of health measures have also been produced(237-240) see Appendix W – Gold standard psychometric properties. The psychometric properties reported in the main development paper for each of the identified instruments were compared with these gold standards (see Appendix W). If the properties had not been reported in the development paper, a further search was undertaken of the papers reporting the use of the instrument to assess whether subsequent authors had commented on the missing psychometric properties.

Results

I identified eleven questionnaires in total, six of which were self-reported. Each of these is described below and a critical evaluation of each is presented in Appendix X – Psychometric properties of SDM instruments and summarised below.

Patient Perceived Involvement in Care Scale (PPICS)

Patient Perceived Involvement in Care Scale (PPICS) (256) is an instrument that measured the degree to which patients feel involved in decision-making. Conceptually, the questions were based on qualitative work that resulted in 25 commonly observed behaviours being defined by the authors during routine primary care visits. These observations were not described and it is not clear why published frameworks were not considered. This instrument had 13 items with two response categories (0=Disagree or 1=Agree) relate to the sub- scales 'doctor facilitation', 'patient doctor information exchange' and 'patient decision making'. Higher scores reflected a greater degree of perceived patient activity and involvement in the medical visit. It was not possible to assess how the questionnaire domains fitted in with the conceptual framework as it was not described. The authors did not report if items reflected relevant domains described, or to the overall construct.

Facilitation of Patient Involvement in Care Scale (FPICS)

Facilitation of Patient Involvement in Care Scale (FPICS) (257) is an instrument that measured the degree to which patients perceive that their personal doctors actively facilitates or encourages them to be involved in their own healthcare. Conceptually, the questions were based on a preliminary pool of 18 statements regarding various aspects of doctor behaviour that may have been observed by the patient over the course of their relationship. Seventeen psychologists reviewed these statements with regards to face validity, content overlap and ambiguity. The resulting critique was used to form the construct of the questionnaire. The original 18 statements or the final construct were not described and it is not clear why published frameworks were not considered. There were 9 items with a 6-point response category (ranging from 0=none of the time to 6=all the time). This instrument had no

subscales and higher scores suggested greater perceived doctor facilitation of patient involvement. The authors did not report if items reflected the overall construct.

Dyadic Observing Patient Involvement Scale (OPTION)

Whilst Observing Patient Involvement scale (OPTION) (253) is an independent observer rated instrument that measured how much the doctor involved the patient in the decision, Dyadic OPTION(248) is an adaptation that allowed patient reported measures of the same underlying construct of doctor competencies required for patient involvement. These competencies were derived from patient focus groups and included: 1. Involvement in problem definition; 2. Explaining equipoise; 3. Communicating options and risks; 4. Conducting the decision process. Based on this construct the Dyadic OPTION was adapted to measure perceived patient involvement from the patient's and doctor's perspective in a single dimension, immediately after the consultation. Three cycles of cognitive debriefing were undertaken to refine the content validity in line with the conceptual framework. The resulting questionnaire had 12 items with a four-point response scale (1= Strongly disagree 4=Strongly agree) with higher scores suggested greater perceived patient involvement in the decision. Whilst the authors undertook an extensive process of ensuring high content validity they did not report how well each item correlated with the overall construct of shared decision making. Only the psychometric evaluation undertaken on the Dyadic OPTION scale are reported in the Appendix X – Psychometric properties of SDM instruments, or critiqued below.

Shared Decision Making 9 Question Instrument (SDM-Q9)

The Shared Decision Making 9 question instrument (SDM-Q9) (258) is a revision of the original SDM-Q, designed to be completed by the patient, immediately after their consultation, that measured the shared decision making process. As a first step towards a theory-driven instrument, the underlying concept of shared decision making (185,259) was refined into nine practical steps (260). Through a process of iteration, the group arrived at nine questions to measure the nine practical steps of their underlying construct. These showed good face validity ratings amongst patients. The resulting questionnaire had nine items with a six-point response scale (1=completely disagree, 6=completely agree), with

higher scores suggesting their doctor had conducted a shared decision making consultation. The 9 items matched the 9 steps identified by the authors in their description of the conceptual framework. However, there was no evidence to suggest that individual items correlated well with the overall framework of shared decision making.

Glass(261) and colleagues tested the English version of the 9-item Shared Decision Making Questionnaire (SDM-Q-9) in a stratified sample of N = 488 respondents in the US. Scholl(262) and colleagues recently adapted the patient-report SDM-Q-9 to a doctor version (SDM-Q-Doc) in order to allow measurement from both view- points (dyadic approach). This scale was tested in medical encounters between 29 doctors and 324 patients in German outpatient care.

Health Care Empowerment Questionnaire (HCEQ)

Health Care Empowerment Questionnaire (HCEQ)(246) is an instrument designed that measured how empowered patients are to take care of their health, in a general sense. The authors described the instrument based on three individual empowerment indicators (feeling of control; interaction with health professionals, and decisional process) but did not cite the literature this comes from. The instrument consisted of ten statements (3 relating to involvement, 3 to control and 4 to interactions) and two response scales. The first response scale described perceptions of how empowered they were (e.g. did you feel that you asked for explanations?) scored on a four-point scale (1=not at all, 4=extremely). The second response scale described how important being empowered was (e.g. how important is asking for explanations) scored on a four-point scale (1=not important at all, 4=extremely important). Each item was scored as a cross product on these two scales, with highest scores indicating the most empowered patients. The individual items reflect the relevant subdomains appropriately, and the subdomains reflect definition of the overall construct of empowerment. The individual items also reflect the overall construct; however, an item-total correlation score was not reported.

The Decisional Conflict Scale (DCS)

The content of Decisional Conflict Scale (DCS) (263) is based on the construct of decisional uncertainty (220), adapted using expert opinion. The DCS has five subscales that have developed following testing between groups and broadly based on the conceptual framework of decisional uncertainty(220,222,223). Subscales include being informed (informed), the decision being reflective of underlying values (values), receiving adequate support for the decision (support), how effective the decision is likely to be (effective), and uncertainty during the decision (uncertainty). The instrument has 16 items scored on a five-point scale (0=Strongly agree, 4= stongly disagree). The total score, and subscores, are transformed to 0-100 scale, with higher scores associated with greater uncertainty and poorer decisions – that is not feeling informed, not feeling the decision reflected your values, not feeling supported through the decision, not feeling like the decision will be effective and feeling uncertain about your choice. The items reflected the subscales described.

The scale discriminated between those who make and those who delay decisions (effect sizes range from 0.4 to 0.8). The instrument had a minimal clinically important difference previously established; patients scoring 25 or more are likely to delay decisions, whereas those with a score of less than 25 tended to make decisions (264). Patients scoring more than 25 out of 100 (265) on the total patient DCS has been related to decisional delay, departure from active treatment, decision regret, nervousness and a higher intention to sue physicians in cases of harms from treatment and has been the score most commonly used to distinguish a harmless from a harmful level of decisional conflict (265-268).

The Provider Decision Process Assessment Instrument (PDPAI (269)) was an adaptation of the DCS that is completed by doctors. The tool included 12 items rated on a 5-point Likert scale that measure a health care provider's decisional conflict with a medical decision. Eight items were taken unchanged from the DCS, one item was a combination of two DCS items (i.e. knowledge regarding benefits and risks) and three new items have been added (i.e. easy to identify all considerations, fully understand the patient's views and satisfaction with the decision-making process).

Psychometric properties

See Appendix X – Psychometric properties of SDM instruments.

Overall, the evidence of psychometric robustness for these six questionnaires was limited, with studies often only quoting one piece of information to substantiate the instrument's reliability and validity. Tests of reliability and validity varied in quality, and information regarding responsiveness was absent in all instruments.

Only three instrument described a process of item reduction (SDM-Q9, HCEQ and DCS), none satisfactorily. Whilst item-total correlation was mentioned by HCEQ and SDM-Q9 authors, missing data and item-item correlations were omitted. Therefore, no instrument reports a thorough evaluation of items.

Acceptability was only reported by authors of the DCS who commented that they had less than one percent missing data in their overall scores. Floor and ceiling effects could be calculated from the data they presented and were less than ten percent. Therefore, the DCS was the only instrument that reported sufficient evidence for acceptability.

Overall, all instruments, except HCEQ, described some of the key processes necessary for content validation. Two of the instruments were based on published conceptual models related to constructs described above (Dyadic OPTION – patient involvement, DCS- decisional uncertainty). Many others had developed their own framework but then refined it through peer-review or cognitive debriefings. SDM-Q, dyadic OPTION and PPICS suggested that their content was based on qualitative work, although the details of analysis were not published.

Internal consistency of all instruments, except the dyadic OPTION, had been reported in the original publications: Cronbach's alpha scores were generally high (0.73-0.94), suggesting that items within each instrument were measuring a similar construct. Therefore, all instruments (except Dyadic OPTION) met minimum standards for internal consistency.

Three instruments had a published test-retest reliability score (FPIS, HCEQ and DCS) and all showed strong correlation scores between test and retest (>0.70). Evidence was not available for Dyadic OPTION, SDM-Q9 and PPICS.

Inter-rater reliability was only appropriate for the Dyadic OPTION scale as it is the only instrument that allowed the measurement of a construct from multiple different perspectives (patient, doctor and independent observer). However, the inter-rater reliability for Dyadic OPTION was somewhat low and below the recommended criterion (R=0.58).

All instruments, except two (Dyadic OPTION and HCEQ), reported some evidence of convergent validity. Correlation coefficients were not always reported, but when they were reported, they were frequently weak associations only (e.g. FPICS to general health perception scale r=0.19). Therefore, whilst commonly reported, the strength of the effect between the instrument being investigated and instruments that measure similar constructs were sub-optimal. DCS was amongst the best performing as its convergence had been tested most extensively and scores showed reasonable associations (e.g. correlation between DCS and perceived risk questionnaire r=0.48)

Discriminant validity was not reported for the HCEQ. For the remainder of the instruments it was reported as lack of correlation between the instrument and certain patient characteristics (age, sex, etc.). No instruments found a relationship between the scale score and these characteristics and therefore, all instruments (except HCEQ) were deemed to have sufficient discriminant validity.

Known groups differences was poorly reported in the primary publications. SDM-Q9 and PPICS were both shown on two sample t-tests to differentiate between 1 and 2 levels of simulated SDM ($p<0.001$). However, this was in a simulated setting, and therefore, may not reflect true differences in shared decision making seen in clinical practice. FPIS failed to differentiate between patients who saw female doctors compared with patients that saw male doctors (the authors described evidence to suggest female doctors involve patients more than their male counterparts). The DCS was the only instrument that has described strong differences in magnitude between groups expected to be different, that is those who chose to undergo screening and those that delayed their decision.

Within scale analysis was undertaken for all instruments except HCEQ and was most commonly tested through exploratory factor analysis. For all tested instruments, except

PPICS, exploratory factor analysis showed expected results, with proposed models describing more than 60% of the total variance.

Responsiveness was only reported for DCS:22 studies (4343 participants), involving decisions about prostate cancer to IVF (270). The meta-analysis showed those randomized to receive a decision aid reported a 7.26 lower mean uninformed sub-score in their DCS compared to those randomized to normal decision management (t-test $p<0.005$).

Summary

Reporting of psychometric evaluations of scales in the field of SDM is inconsistent. It could be improved and be more consistent by adhering to the now well established psychometric criteria advocated by bodies such as MOT, COSMIN, and FDA. Many of the presented instruments show satisfactory, good or even excellent internal consistency, but other measures of reliability such as test-retest reliability have not been reported. Validity has in several cases not been sufficiently tested, and reported evidence is sometimes weak. Responsiveness was not reported in any instrument other than the DCS.

From the available instruments the DCS appears to be the most psychometrically robust, and measures a construct that is very pertinent to surgical rate variation, as discussed in the introduction: Decisional conflict scale allows for assessment of decisional uncertainty from both the doctors' (using the PDPAI) and patients' perspective, and allows a more thorough investigation of the decision-making process. Additionally, surgeons' decision uncertainty may be a surrogate marker for professional uncertainty, or at least influenced by it. DCS, therefore, it is a strong candidate as an instrument to investigate the drivers of surgical rate variation, as they manifest during the consultation.

Conclusion

There are many definitions of SDM, with seven main underlying overlapping concepts. Whilst there is some evidence that professional uncertainty also has a role regional surgical rate variations this has never been measured at the consultation level, and it is plausible that professional uncertainty leads to decisional uncertainty of both the patient and doctor.

Therefore, decisional uncertainty was considered a valid concept to measure in a study of regional surgical rate variations. The DCS, as a measure of decisional uncertainty, has reasonable evidence of reliability, validity and responsiveness, and is therefore the best available questionnaire both conceptually and psychometrically for use in our study on regional adult tonsillectomy rate variation. The following chapter will describe measures of preference followed by design of an instrument that can elicit treatment preferences for adults with recurring tonsillitis. The study using both instruments is described in Chapter 7.

CHAPTER SIX: Development of a new instrument to measure patient treatment preferences in recurring tonsillitis

Chapter synopsis

In this chapter I have introduced the standard methods of measuring preferences and justified the use of the Analytical Hierarchical Process (AHP). I have summarised the development of a new AHP based instrument to elicit treatment preference in adults with recurring tonsillitis. This included a systematic review and critical appraisal of existing studies of outcomes for adult tonsillectomy, thematic analysis of new qualitative data about how patients make judgements about these outcomes, expert-panel review to arrive at a final set of outcomes (reducing days of sore throat, reducing halitosis, reducing visits to the GP, reducing risk of bleeding, improving quality of life) for inclusion in the treatment preference instrument and piloting of the instrument prior to data collection.

Introduction

Treatment preferences (of both patient and surgeon) have been shown (see chapters 1) as potentially important constructs that may contribute to regional surgical rate variation.

Based on a seminal cross sectional study of paediatric tonsillectomy rates by region(16) and qualitative study of recurring tonsillitis consultations by region (17), Wennberg (271) hypothesised that regional tonsillectomy rate variation resulted from regionally aligned surgeon preference, a theory he coined the "Surgical Signature". If patients who see surgeon A are more likely to receive tonsillectomy than if they saw surgeon B, it can be hypothesised that surgeon A has a treatment preference for tonsillectomy. However, in the modern health care setting of shared decision making, where the surgeon incorporates the patients' preferences, with the best available evidence, to help reach a treatment decision rather than paternalistically dictating the treatment, surgeon preference is a complicated construct. Consultations with surgeon A could have resulted in tonsillectomy being chosen more frequently because the surgeon perceived tonsillectomy was a more effective treatment for certain outcomes, perceived patients prioritised these outcomes more than others, proxy bias, or the surgeon's personal preference towards tonsillectomy. Partly due to the complexity of this construct surgeon treatment preferences, as they apply in the medical consultation, have always been difficult to measure on a scale large enough to inform discussions about regional surgical rate variations. Instead they have been implied from smaller qualitative studies, especially around the surgical management of breast cancer(26). There is no evidence that directly reports on surgeons' treatment preferences for the management of adults with recurring tonsillitis, and how this affects treatment decisions.

In relation to regional surgical rate variation, whilst surgeon treatment preference has a historical context and has been discussed (disproportionately in relation to the level of current evidence) patient treatment preferences have only recently come to the fore. There is now a growing body of evidence from studies of orthopaedic conditions that regional surgical rate variations are associated with regionally aligned patient treatment preferences (78). (78,81-83) (85-88), (272) (90) (91). However, due to the difficulties in eliciting patient treatment preferences efficiently on a large scale, preferences have been implied by willingness for treatment based on selecting an ordinal value on a spectrum from "*definitely*

not willing to have to treatment", through to "*definitely willing to have treatment"*, has never been tested with respect to validity or reliability. There is no evidence that directly reports on treatment preferences for adults with recurring tonsillitis, and how these patient preferences influence the treatment decision.

The formation of treatment preferences is a complicated process involving the combination of an evaluation of the intervention and the choice between alternative interventions based on that evaluation. However, as complexity of information increases so does the probability that the respondent departs from the rational decision making and employs simplifying heuristics to form a preference, and therefore, omits the evaluation phase(234). Therefore, it is important that preferences be measured in a reproducible and meaningful manner that also takes account of simplifying heuristics. Unfortunately, most preference elicitation methods have either not accounted for differing methods of preference generation (rational decision versus heuristic decision) or had high cognitive burden for the respondent.

To undertake a study that allows us to investigate the role of treatment preference for adults with recurring tonsillitis it is important to have an instrument that can accurately elicit treatment preference. Therefore, this chapter describes the development of an instrument designed to quantify treatment preferences in adults with recurring tonsillitis. Initially, I evaluated common methods of preference elicitation (i.e. Standard Gamble, Time-Trade-Off, Visual Analogue Scales, Conjoint Analysis, Analytical Hierarchical Process) and provided rationale for why I selected the Analytical Hierarchical Process for our study. Thereafter, this chapter focuses on the development of an instrument designed to elicit treatment preferences in adults with recurring tonsillitis based on the Analytical Hierarchical Process (AHP).

Methods of preference elicitation

The theoretical framework for preference elicitation tends to come from either an economic perspective (based on utility theory) or a psychological perspective (based on expectancy value theory). The economic perspective and utility theory is based on the four axioms of the rational decision maker described by Von Neumann (233). These axioms are: 1. Completeness (assumes that an individual has well defined preferences and can always decide between any

two alternatives); 2.Transitivity (assumes that an individual always decides consistently); 3. Independence (assumes the order of preference cannot be changed by adding an irrelevant option); and 4. Continuity (assumes that preference ordering can continue for a new third option, based on its preference proximity to a previously preference-defined option). Methods based on this theoretical perspective include Standard gamble (SG), Time-trade-off (TTO) and Visual Analogue Scales (VAS). These methods usually aim to obtain consensus amongst a large pool of respondents to identify where treatments (or health states) sit on a preference scale. It is hypothesised that preferences elicited using one of these methods could be compared across all contexts and studies that used the same method.

In contrast, expectancy value theory, based on the work of Fishbein & Ajzen (194) suggests that people develop a belief about the action (e.g. that a particular treatment is painful). The belief can be modified if new information becomes available. People then assign a value to that belief (e.g. that pain is a negative attribute and should be avoided) and then an expectation is created (or modified) based on the combination of the belief and the value (e.g. that the treatment is not a good idea). In psychology, the concept of preference is similar to the idea of attitudes (a disposition of a person to respond favourably or unfavourably to an object, person, institution or event(232)). The two most commonly used methods, based on expectation-value theory, are conjoint analysis, CA, which frequently uses the discrete choice experiments (DCE) format,(273) and Analytical Hierarchical Process AHP (274-277) and studies using these methods for medical purposes has grown substantially in the last decade(278). The difference in these techniques relates to how the respondent is presented information for comparison. CA uses a holistic approach, presenting all information for a single comparison, whereas AHP uses a decomposed method, breaking down the treatment into simpler pairwise comparisons. Methods of preference elicitation developed using expectancy-value theory tend to indirectly measure preferences and then weight the result to transform the scores into ranked preferences. These methods do not require consensus amongst respondents, rather the ability to discriminate between them.

Comparison between Utility Theory and Expectancy-Value Theory

In contrast to utility theory, expectancy-value theory allows comparisons across all outcomes related to treatment choice. For example, utility theory would allow comparison of health

outcomes based on treatments chosen, but would not incorporate differences in the process, such as costs, travel times, invasiveness nature of the treatment(279-281). Therefore, the expectancy-value theory describes a more complete model of patient treatment preference. It must be noted that both theories have failed to consistently predict decision-making behaviour and a newer theory of decision making has now emerged that it is argued to more accurately describe how people make their decisions – Prospect theory (282). Prospect theory is based on the original expectancy value model; it predicts that individuals tend to be risk averse in a domain of gains, or when things are going well, and relatively risk seeking in a domain of losses, or when things are not going well. Unfortunately, there are no methods of preference elicitation based on prospect theory yet.

Common methods of eliciting patient treatment preferences

Standard Gamble (SG)

SG has been considered the gold standard method for preference measurement. SG is a direct measure of preference that requires respondents to choose between two options: The first option has a known outcome i.e. the respondent would carry on with their condition the way it is for the rest of their life. The second option includes a gamble: a treatment that has p chance of restoring them to complete health and 1-p risk of death. By systematically varying the risk of death, a threshold can be calculated that standardises all treatments onto the same scale.

This technique requires respondents to weigh up the probability of the two possibilities and come to a decision based on abstract information. This is a complex task that has been shown to require a large amount of cognitive effort (283). Partly due to high cognitive burden, SG cannot be self-completed requires an interview format to collect the data, it is an expensive and time -consuming method to deploy on a large scale.

The SG uses a ratio scale with death and perfect health as their rational zero or ones, respectively and therefore has the advantage that it remains invariant over all transformations where the scale is multiplied by a constant. Direct measures assume that respondents are internally capable of generating an interval or ratio scale. This means that the scale remains essentially the same when it is expressed in different units (e.g. feet rather

than inches). However, there is evidence that suggests that death may not be an appropriate rational zero, since there are some conditions that are considered by some people to be worse than death (e.g. constant severe pain or a vegetative state)(284), thus making it difficult for respondents to use this method. It has been suggested that error associated with these difficulties can be reduced by averaging judgments over respondents(285).

Time trade-off (TTO)

TTO was developed as a simple to administer alternative to the standard gamble and involves a choice between two certain outcomes (286). The respondent is asked to choose between continuing with their condition until death or a treatment that gives them perfect health for *x* years and then death. The interviewer can vary the duration of perfect health until the respondent can no longer differentiate between options. Since TTO cannot be self-completed (i.e. it relies on an interviewer to administer questions) and is most commonly completed on a ratio scale it has similar limitations to SG (i.e time consuming, expensive to deploy in studies and assumes respondents can generate preferences on a ratio scale.)

Visual Analogue scale (VAS)

Respondents are asked to place their mark of preference for a treatment on a pre-labelled scale anchored at either end with markers of best and worst health imaginable (e.g. perfect health and death, respectively). VAS provides interval level information and therefore indicates how far apart the treatments are as well as their rank order, but it does not indicate the absolute magnitude of preference for any treatment. Another potential disadvantage of having only an interval scale is that algebraic operations can only be performed on intervals and not on scale values, so it cannot be said, for example, that one health state is twice as desirable as another. However, for most practical purposes an interval scale is sufficient. Additionally, there is considerable evidence now that shows visual analogue scales for a concept as complicated as preference for treatment may lack validity(287,288), partly because human cognitive capacity is unlikely to be able to make judgements on a 0-100 scale(289-291).

Conjoint Analysis (CA)

CA asks patients to compare clinical vignettes of potential outcomes. CA elicits treatment preference indirectly onto an ordinal scale by asking respondents to make comparisons between treatments, in a stepwise fashion, so that the overall ranking of all treatments can be calculated. Since each vignette is slightly different from the last, it allows the investigator to calculate the decision-maker's individual priorities(273,292,293). CA approximates a real life decision, where the integration of information is down to the respondent, and has been shown to imitate the respondents judgement in live decision-making(294). However, it has high cognitive burden related to high volume of information, and need to remember information to undertake task effectively (295). The ordinal scale is the most primitive form of measurement, and requires the respondent to rank order a set of objects (treatments). In addition, there is no indication of how much preference each treatment possesses or how far apart the treatments are in terms of preference. Ordinal scales provide limited information and none of the fundamental operations of algebra may be applied(296).

Analytical hierarchical process (AHP)

AHP decomposes treatment into potential outcomes and asks patients to make pairwise ratio-comparisons between potential treatment outcomes. Ratios of preference between different outcomes are combined with probabilities that outcomes are likely to occur if each treatment was chosen in turn to calculate a treatment preference score(274). Therefore, AHP indirectly places preferences onto a ratio-scale. Pairwise comparisons required of AHP are less cognitively taxing than CA since they involve less information per step and are less time consuming compared with composite comparisons of conjoint analysis(297). But it has been argued that pairwise comparisons are arbitrary differences in factors that do not model the decision making problem(298). AHP can also quantify how inconsistent respondents are with regards to the axioms of rational decision making.

Comparison of methods

Whilst instruments of preference elicitation have not been formally psychometrically evaluated, existing evidence suggests that people exhibit patterns of preference that are often incompatible with expected utility theory. For example, a study found (299) that after changing the outcomes being compared in the standard gamble technique (e.g. from death

to severe disability), respondents' evaluations of treatments became unstable, a finding that both contradicts expected utility theory and indicates that the standard gamble is internally inconsistent. There is now extensive evidence that people violate utility theory when making live decisions(300). The idea that tools based on utility theory may lack validity is further seen in studies that compare SG, TTO and scaling methods, most of which find no convergent validity (287). In using these methods, economists have been chiefly concerned with defining the preference location of treatments and diseases on a standardised scale and so require large numbers of people to undergo measurement. Errors in measurement of preference created at the individual level were considered to be reduced when population aggregates were taken. Evidence suggests that AHP performs better than CA(294,297,301,302). In addition, AHP generally relies on the axioms of rational decision-making but allows the decision maker to be inconsistent and intransitive to some extent, thereby taking into account the difficulty of giving precise preference judgements.

Which measurement method is best for adults choosing a treatment for their recurring tonsillitis?

Currently, there is no ideal method to elicit treatment preference and no instrument that elicits preference with regards to treatment choice of recurring tonsillitis, whether based on utility or expectation-value theories. An instrument that elicits preference for adult tonsillectomy using the AHP would be advantageous for three main reasons.

AHP is the only method that not only allows for, but also measures, violations in the axioms of rational decision making. Kahneman and Tversky have shown that the axioms of rational decision making are frequently broken in people making decisions and so methods of measurement based on these axioms fail to validly describe preference (282).

The AHP method has amongst the lowest of cognitive burdens, due to the simplicity of undertaking pairwise preference comparisons and is preferred by respondents especially when there are more than six outcomes to compare. As a result of relatively low cognitive burden, the AHP can be easily deployed as a questionnaire with minimal costs and use of manpower required for interviews of SG or TTO. The AHP approach has been well validated(303) and has been shown to have high predictive validity when compared to conjoint analysis (294,301).

AHP preference generation process

AHP requires participants to undertake a series of pairwise comparisons between all potential treatment outcomes, based on their personal priorities (that is 7 treatment outcomes would require 21 pairwise comparisons). Comparisons are made on a response scale with semantic descriptors (e.g. Equal preference, Moderate preference, Strong Preference, Very Strong Preference, Extreme Preference) and later converted to the Saaty scale so that equal preference between two outcomes would score: 1, Moderate preference; 3, Strong preference;– 5, Very strong preference;- 7 and Extreme preference. - Pairwise scorings are entered onto a Saaty matrix (304) which has four final outputs (process described Appendix T AHP methodology and Ranking results): Ranked outcome priorities, Outcome preference score and consistency score, Treatment preference score

Ranked outcome priorities are an ordinal rank that relates the preference of outcomes to each other. Outcomes with higher ranks show higher preference. Whilst they show the order of preference they do not provide the magnitude of preference between outcomes.

Outcome preference scores provide a value of preference for each outcome on a ratio scale. That is the magnitude of difference between outcomes can be calculated using these scores.

Consistency scores display how consistently participants completed all their comparisons. For example, if a participant initially stated that he/she preferred A twice as much as B, and B twice as much as C, then he/she should prefer A four times as much as C. Gross deviations from this are measured in the AHP consistency score. This score is an internal validity check of the process. Scores greater than 0.2 show inconsistent responses, and score greater than 0.5 equate to random responses.

Treatment preference score is a weighted score that shows how much their outcome preferences line up with a treatment. If a participant has a strong preference for a treatment that reduces halitosis, but the evidence suggests that Treatment A has no effect on halitosis, then his preference for halitosis will be down-weighted. Alternatively, if a participant has a moderate preference for a treatment that reduces sore throat days and the evidence shows that Treatment A has a large effect on reducing sore throat days, his preference for reducing sore throat days will be up-weighted. Scores are normalised and added together to provide

an overall preference towards Treatment A. If the choice of treatment was between Treatment A and Treatment B, and the AHP provided a Treatment A preference of 0.5, it would mean that the participant had equal preference for treatments.

The remainder of this chapter is dedicated to describing how I developed an instrument that measures treatment preference for recurring tonsillitis using the AHP approach.

Development of Preferences in Adults with Recurring Tonsillitis Tool (PARTTS) -- Overview

See Figure 17 Development of the PARTT.

To develop a new instrument measuring preferences for the treatment of recurring tonsillitis I needed to: 1) identify the key outcomes relevant to management of recurring tonsillitis; 2) ascertain the relative importance of these outcomes; 3) shortlist the potential outcomes to identify a manageable number (Figure 17). I undertook a **systematic review** and critical **appraisal** of randomised and non-randomised studies that described outcomes in relation to the treatment of recurring tonsillitis. To ensure that these outcomes were important to patients making their treatment decisions I presented all identified outcomes to relevant patients who were asked to rank them in terms of importance –**patient ranking of treatment outcomes.** To confirm their rankings a further sample of these patients were invited to a **focus group** to discuss why particular outcomes were considered important. All outcomes were considered by an **expert panel** to identify a manageable number of shortlisted outcomes in the final instrument (seven or fewer). Outcomes were chosen based on their importance to recurring tonsillitis patients and the reliability of evidence on their expected likelihood of occurrence after treatment. These final set of outcomes were developed into a questionnaire using the AHP format. The instrument was piloted on patients with recurring tonsillitis - **Pilot**. The following subsections will describe the Methods and Results for each of these stages in turn.

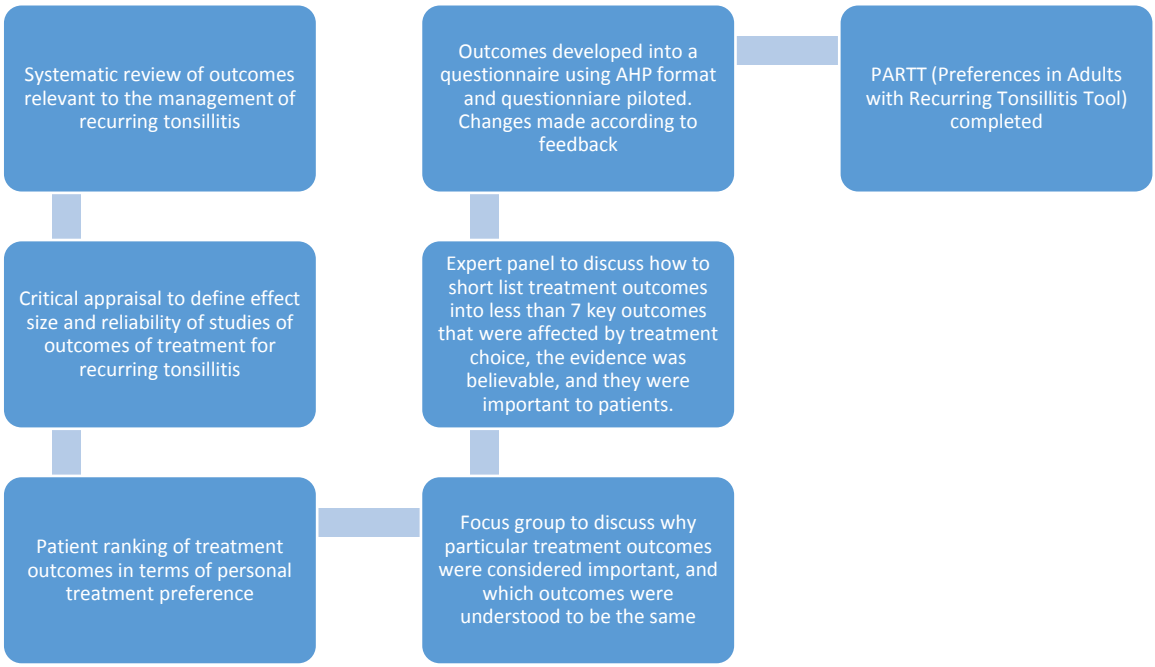
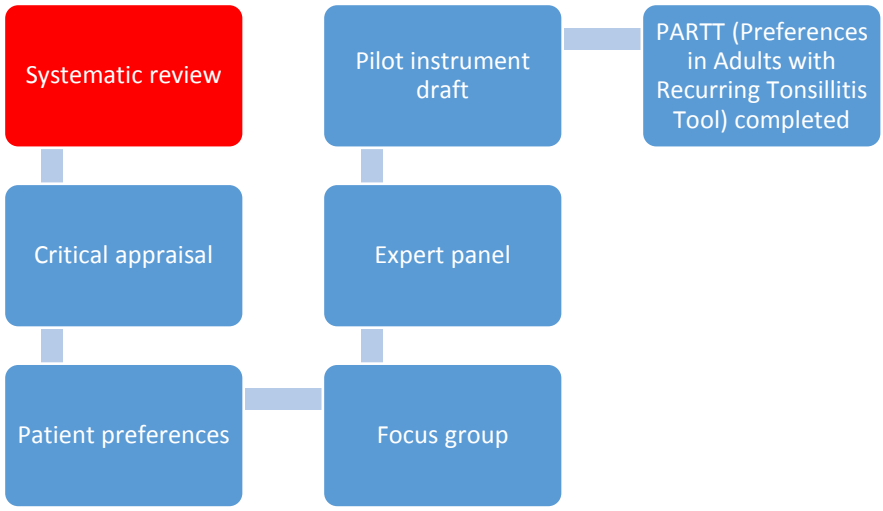


Figure 17 Development of the PARTT

This shows all the stages involved in the development of PARTT, allowing evaluation of treatment preferences for adults with tonsillitis.

Identifying Outcomes (systematic review & critical appraisal)



Method

Search strategy

I searched the following databases from their inception for published studies that reported medium-to-long term outcomes in tonsillectomy undertaken for adults with recurring tonsillitis: PubMed, EMBASE and CINAHL. I modelled subject strategies for databases on the search strategy by a Cochrane systematic review looking at the effectiveness of tonsillectomy (see Appendix E Search strategy for tonsillectomy outcomes), since our study was to investigate how patients choose between tonsillectomy and watchful waiting in the treatment of recurring tonsillitis.

Inclusion criteria

- Adult tonsillectomy for recurring sore throat/tonsillitis
- Medium-Long term outcomes
- Randomised and non-randomised studies

Exclusion criteria

- Paediatric tonsillectomy
- Tonsillectomy undertaken for cancer
- Outcomes that are intraoperative (e.g. blood loss)
- Outcomes that are perioperative (e.g. post-operative nauseas and vomiting)
- Outcomes related to the histopathological or microbiological characteristics of the excised tonsils
- Not in the English language

Data extraction

Two researchers (NM & LM) independently extracted data from the included studies using standardised forms.

Selection of studies

Studies were grouped by outcome, i.e. studies reporting on quality of life were grouped with other studies relating to quality of life. A study was placed in more than one category if it reported more than one relevant outcome. Within outcome groups, studies were excluded if

there were higher level studies already in the group (e.g. a case report on quality of life was excluded if there was also a systematic review reporting on quality of life after tonsillectomy).

Three researchers (NM, LM, OR) independently screened titles and abstracts obtained from the database searches at different stages of the original review and subsequent updates. Similarly, at least two of the three researchers (NM and LM) independently reviewed the full text of the potentially relevant titles and abstracts against the inclusion and exclusion criteria. Studies were classed as Definite Include, Probably Include, Probably Exclude and Definitely Exclude. Studies that were classed as Probably Include or Probably Exclude were discussed to decide whether they would be included. Differences in classification were also resolved by discussion.

Remaining studies were critically appraised using the appropriate CASP tools (305-308) and effect size and study quality of cross-tabulated for each outcome.

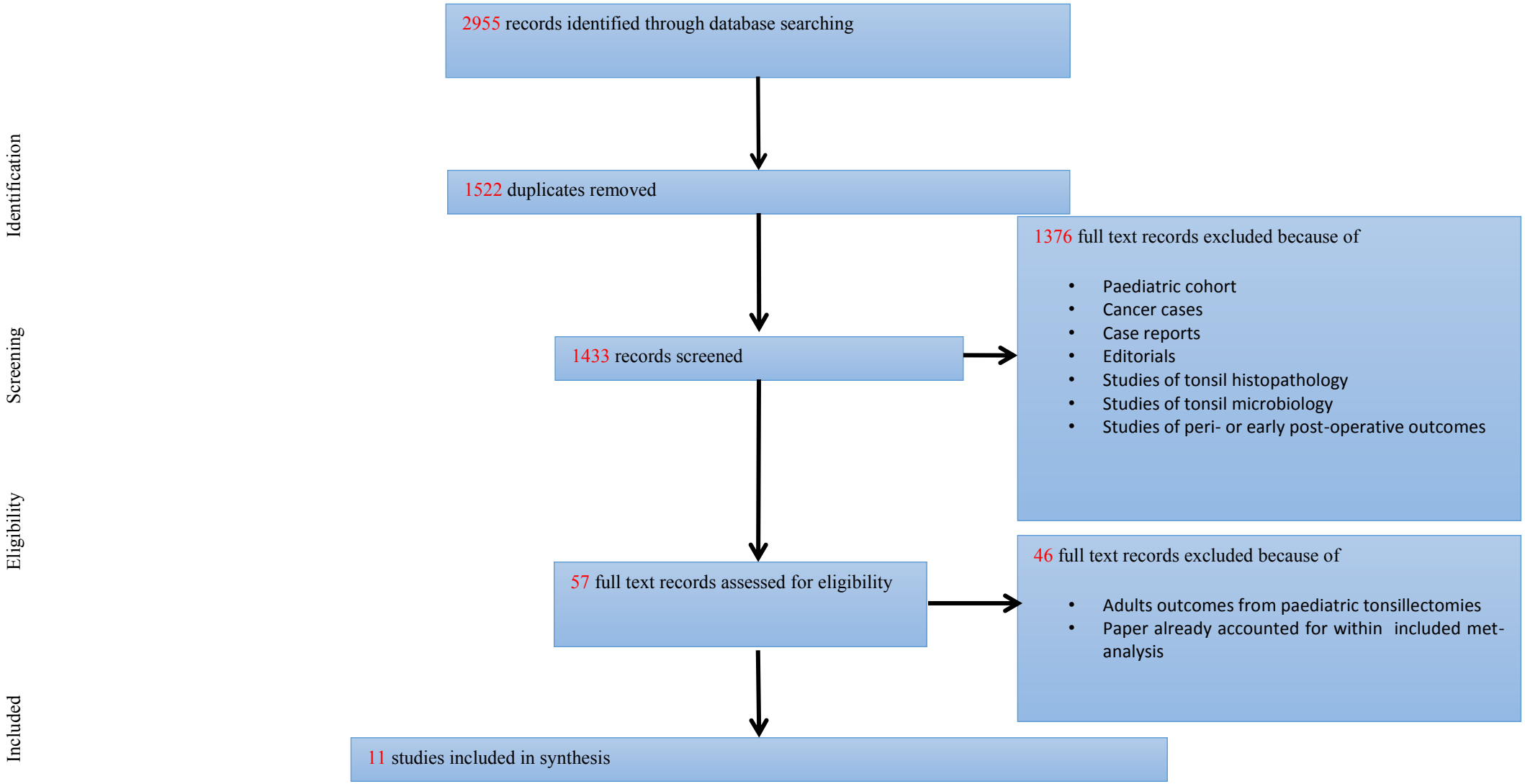


Figure 18 Systematic review of literature for treatment outcomes related to recurring sore throat

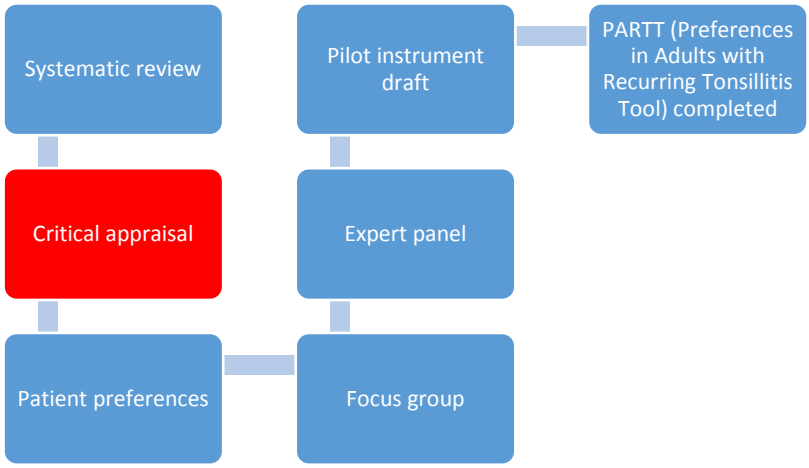
This flow chart shows all the stages involved in creating the systematic review used to derive possible treatment outcomes related to recurrent sore throat.

Results

See Figure 18 Systematic review of literature for treatment outcomes related to recurring sore throat.

The initial search returned 2955 abstracts. Removal of duplicates yielded 1433 abstracts. Removal of paediatric cases left 870 abstracts. Removing tonsillectomy for cancer and case reports left 175 studies. Removal of editorials and studies that pertain histopathology or microbiology left 57 studies. Finally, studies that reported on adult outcomes on patients who had undergone a tonsillectomy as a child were removed to leave 19 studies.

If more than one study reported on the same outcome (e.g. reducing days of sore throat) only the study with the highest level of evidence was included and the others excluded (e.g. meta-analysis of RCTs examining reducing days of sore throat retained by individual RCTs excluded). Exclusion based on highest levels of evidence in outcome groups produced eleven studies that had measured different outcomes of tonsillectomy: days of sore-throat; the number of episodes of sore-throat, visits to the GP, voice, haemorrhage, snoring, halitosis, taste disturbance, alteration to immune profile, quality of life and societal cost.



Whilst I had identified all available tonsillectomy outcomes that had been published, the highest level of publication related to those eleven outcomes, I was still not clear how believable (reliable, valid) the results of these publications were, or how meaningful they were to this study (i.e. effect size and generalisability).

Therefore, I undertook a critical appraisal to assess the risk of bias for each publication. Since the publications ranged from meta-analyses of RCTs through to case series I used the

Critical Appraisal Skills Programme (CASP) risk of bias checklists. These checklists were developed to detect risk of bias in RCTs, cohort, case-control, case-series and qualitative studies. CASP have provided a robust framework for reporting studies by forcing to appraiser to answer 3 key questions: 1. Are the results valid? 2.What are the results? 3. Do the results apply to your population?

To address question number 2 and allow comparisons across studies I calculated standardised effect sizes of tonsillectomy on the outcomes (difference in means between those who received tonsillectomy and those that did not/pooled standard deviation). The effect size of tonsillectomy on each of the eleven outcomes was categorised into small (0.2-0.49), medium (0.5-0.79) and large (>0.8) (309). When insufficient data was presented to make this calculation (e.g. immunological profile) the results were discussed with two ENT surgeons to help categorise the impact. The study was graded based on Oxford Clinical Evidence Based Medicine Levels of Evidence(310).

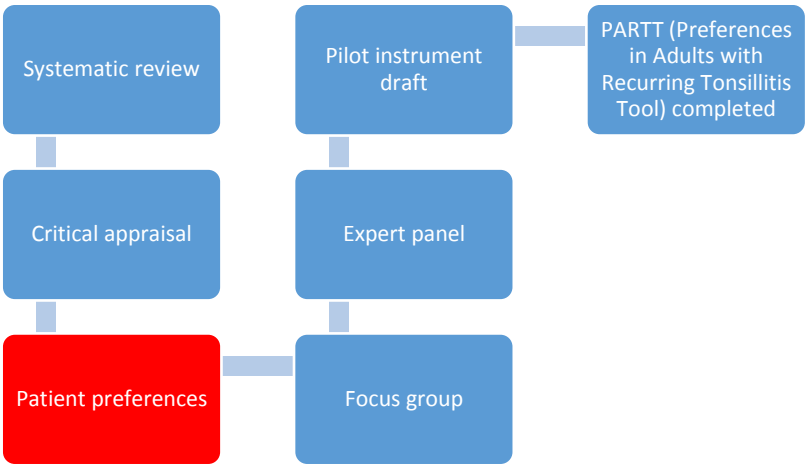
Based on the critical appraisal of these eleven publications I graded the evidence with regards to the outcomes as Strong, Moderate and Weak. **Weak** was classified as any study that was level 3 (case-control), 4(Case series) and 5 (expert opinion). **Strong** was classified for any study that had level 1 (randomised controlled trial) or 2 (cohort) evidence and reported a large effect size (Cohen's D>0.8). **Moderate** described all other combinations.

For the full critical appraisal of eleven studies that allowed us to classify the effect size into small medium or large, or the strength of study into weak, moderate or strong please see Appendix F – Critical appraisal of tonsillectomy outcome studies using CASP and Summary.

Conclusions

I identified eleven outcomes of which four had strong evidence (days of sore throat; number of episodes; visits to GP; risk of haemorrhage) as outcomes of tonsillectomy. Evidence relating to other outcomes was weaker. It is not clear if any of these outcomes were important to the patients making their treatment decisions. Therefore, the next stage was to examine how patients with recurring tonsillitis valued these outcomes.

Patient ranking of treatment outcomes (questionnaire ranking and focus group)



Methods

Ethical approval for this study was granted by Nottingham Research Ethics Committee (15/EM/0191) on 8th May 2015.

I contacted, by email, all patients who had attended ENT clinics with recurring tonsillitis at the Royal National Throat Nose and Ear Hospital from 2010 to 2011 and agreed to be part of a research register (n=35). Twenty-three patients had previously chosen to treat their condition with tonsillectomy, whilst twelve had chosen a non-surgical pathway. (watchful waiting). Patients were asked if they would be willing to participate in an online ranking exercise and/or attend a focus group to discuss which factors had been important to their treatment decision.

Twenty-six patients (of the original 35 contacted) agreed to undertake the online ranking exercise and were emailed the questionnaire (Appendix G – Online ranking exercise). The questionnaire had a description of the 11 outcomes shortlisted from the systematic review and a brief description of what each one meant. They were asked to rank the eleven outcomes in order of importance to their treatment decision, with number one being the most important and number eleven the least. They could rank items with the same number if they felt outcomes were equally important.

To develop a better understanding of the importance of the eleven identified outcomes when patients make decisions about treatment for recurring tonsillitis I conducted a follow up focus group to address the following questions:

Which factors drive patients' treatment decisions when they suffer from recurring episodes of sore throat?

What process do respondents go through to rank factors relevant to the treatment of recurring sore throat in terms of importance in their decision?

How well do participants understand terms used to describe evidence based factors relevant to the treatment of recurring sore throat?

The participants were purposively sampled to include those who had chosen to treat their condition with a tonsillectomy and those who had not. From the twenty-six participants who had agreed to take part in the online exercise, eight agreed to attend two planned focus group meetings. All eight patients who had agreed were emailed an information sheet about the focus group (Appendix H - Patient Focus group information sheet). Five of the eight patients attended on the day and were consented (Appendix I – Consent for focus group) to participate in an audio recorded focus group. This meant there was one focus group with four attendees (2 who had chosen tonsillectomy and 2 that had chosen watchful waiting) and another focus group with only one (who had chosen tonsillectomy). The latter was turned into a semi-structured interview.

Two interviewers were present for the focus group and semi-structured interview (NM and AD). A semi-structured discussion guide was drawn up (Appendix J – Semi-structured interview guide) prior to the meeting to ensure the above three objectives were covered.

The interviews were audio-recorded, anonymised and later transcribed.

During the focus group, I asked participants to rank the eleven evidence-based outcomes in order of personal importance to their treatment decision. After they had completed the exercise, I asked the focus group participants to discuss the process of how they had ranked the factors. After a period of discussion, I gave the participants further information about the

outcomes (such as the likelihood of outcome) and then asked them to repeat the ranking exercise.

Content analysis of each transcript was conducted manually. Two researchers read each transcript and independently identified the main themes line by line. Discrepancies within these themes were discussed between the researchers until consensus was achieved. Discussions led to certain themes being combined whilst others were rephrased. The emerging conceptual framework was reviewed by members of the supervisory team who are from a multidisciplinary background and is described below.

Results

The online ranks given by participants are shown in Table 24. Respondents ranked improving the quality of life as the most important factor in their treatment decision. Reducing days and number of sore throats and visits to the GP for a sore throat followed closely behind improving the quality of life. Reducing halitosis was the fifth most important outcome to our respondents. Respondents ranked reducing the risk of haemorrhage and taste disturbance as joint sixth outcomes. Reducing snoring was ranked eighth and reducing the risk of damaging the immune system ranked ninth. Reducing cost of treatment to society and improving voice were ranked lowest.

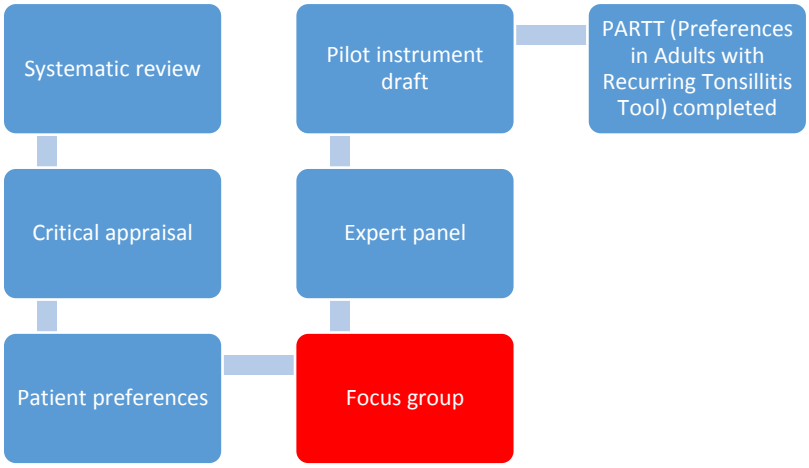
List of factors	Average patient rank	Rank
Reducing days of sore throat	3.6	3
Reducing number of episodes of sore throat	2.2	2
Reducing visits to the GP for sore throat	4.6	4
Improving quality of life	1.8	1
Reducing snoring	5.2	8
Improving voice	10.6	11
Reducing halitosis	4	5
Reducing cost of treatment for society	8.2	10

Reducing risk of taste disturbance	7	6
Reducing risk of haemorrhage	7	6
Reducing the risk of damaging the immune system	8.4	9

Table 24 Patients' ranked preference of potential treatment outcomes.

This table shows the patients' ranked preferences in relation to the treatment outcomes for recurring sore throat. The most important three were improving quality of life, followed by reducing the number of episodes of sore throat and the number of days of sore throat. n= 26 patients

Which factors drive patients' treatment decisions when they suffer from recurring episodes of a sore throat?



Thematic analysis of the transcriptions generated four conceptual domains:

1. Negative aspects of the disease
2. Hope for improving current situation
3. Knowledge of potential treatments
4. Individual trade-off between risks and benefits

Theme 1: Negative impacts of the condition

Many participants talked about their decision to have treatment as motivated by the negative impact of the disease. This included the intensity of physical symptoms and the psychosocial impact of symptoms.

Physical symptoms

For some, the intensity of physical symptoms was described in terms of the duration of time over which episodes had occurred, "Oh, a good couple of years but it kind of got really bad the end of my first year of training, second year, going into my second year". Other participants described the frequency of occurrences "Since I came to this country it was eight times since last October. I had tonsillitis eight times!" Others described the intensity of physical symptoms in terms of their perceived severity, using phrases such as "really bad" and "awful". The sense of deterioration was described by most participants and the phrase "getting worse" was frequently used.

Some participants described their susceptibility to colds as a negative impact of their condition "As in like I had it [tonsillitis] a lot, so whenever I had tonsillitis, I have a rubbish immune system anyway, but it would be worse, like I'd be so ill and I'd easily get any other colds or anything…"

Many descriptions focussed on the negative impact of other troublesome symptoms in addition to sore throat ("snoring", "difficulty breathing", "sleep", "bad breath", "drooling") on themselves, but often also included the impact of symptoms on others, "...I was snoring which affected me like my girlfriend was complaining about it…" or "…Every time I'm sick I'm eating tons of chewing gum. I'm shy speaking to people because I think that my breath is so bad that no one will talk with me".

Psychosocial impact of symptoms

Later discussions included the psychosocial effect of these symptoms on patients' everyday lives. This included an impact on both professional and personal lives. Participants were concerned about the numbers of days off that resulted from tonsillitis, which resulted in the loss of professional opportunity, for example, "... I sing, act and dance so while I was in my

training when I was getting constant sore throats and tonsillitis, I couldn't do singing exams, I couldn't…where there were opportunities, producers or casting directors or people that would come to the school to hear us sing or perform". Others described this in relation to poor concentration: "I had to be focused and I couldn't be because I was thinking only about my throat and how I'm feeling miserable."

Participants also described the way that others did not understand "If someone is not suffering from tonsillitis, they will often ask me and they won't understand." The effect of this was described in terms of feeling "miserable". This effect on participant's emotional state was also seen in relation to social isolation, for example: "….I've lost contact with my friends because they were going out and I wasn't able to do everything I want to do, I was thinking in terms of tonsillitis…"

Theme 2: Hope for a positive outcome

All participants discussed hope during their conversations. Hope for a positive outcome was nearly always discussed in the context of loss of hope for spontaneous resolution and the need for finality or and ending.

Nearly all participants agreed that they had lost hope of spontaneous resolution "...they've been so persistent since 2008 [when] I first noticed them, they're not going to go away." All participants discussed the need for finality: "I want them gone", some with a sense of anticipation. Most participants talked about the hope of a positive outcome: "It has to work, it just has to. It can't possibly get any worse anyway." Participants described the positive outcome in terms of relief of symptoms such as drooling and snoring.

Theme 3: Perceived availability of treatment options

Knowledge of available treatments influences most participants. Some actively sought this information from their doctors: "...I went to the doctor and asked if there was something that she can do about that and she proposed that she can send a note for tonsillectomy, so I was relieved". Some sought information actively from the internet: "I looked up a lot online what I was going to do…" Whilst by others information was passively received "..but as I know, because I see some people who took out their tonsils but at least two or three, one of my brothers also took out his tonsils…".

Theme 4: Individual trade-off between risks and benefits

In the context of receiving information on the availability of treatments participants also discussed the individual trade-off between short term risks and long term outcomes "When I was thinking about tonsillectomy, I thought that moment of suffering after surgery, it's better than suffering from tonsillitis over and over again, so…"

Whilst another participant acknowledged that the trade-off is different between patients and depends on their disease severity: "I guess when you're deciding it's how much they affect you, because it can just range from being annoying to it actually affecting your day to day life, so I guess it's where you fall in that spectrum, in that scale, and I think all of us are different in different levels, so it would affect us differently, which is why maybe some might not want it and some, yes…"

How do participants rank evidence-based factors relevant to the treatment of recurring sore throat in terms of importance in their decision?

Most participants felt that receiving extra information about the factors made the ranking exercise easier: "It was very helpful for me. Now at least I have some information about tonsillitis before the operation and after the operation and that is very helpful for this."

However, some felt it made it harder. When asked about whether having extra information helped with the ranking exercise one participant responded: "I think that all the vague things, the general things, aren't really of help for this [exercise]."

All participants employed multiple strategies to rank the factors. I coded ranking strategies into four broad conceptual domains:

1. Prioritizing important factors
2. Trading factors off against each other
3. Deprioritizing less important factors
4. Clustering factors based on perceived similarities

Theme 1 Prioritizing important factors

Nearly all participants described starting their ranking exercise by prioritizing the most important factors: “There were some things that were immediately more important to me that I could put as my one and two but then I kind of got lost towards the middle and the end.” Prioritization was most frequently discussed in terms of self over society: “I think that reducing financial cost to society; it’s mostly irrelevant for most people. It wouldn’t change my mind about tonsillectomy.” People frequently and most easily prioritized the negative impact of their disease (defined above) over societal benefit citing the personal importance of their physical health and quality of life.

Some participants discussed prioritization in the context of their future health: “My first choice is reducing chance of altering the immune system, that was the main reason I wanted the tonsillectomy, because I am suffering from tonsillitis very often and I’m taking a lot of antibiotics and they are good because they are treating me but they also make harm to my body and when I was thinking about the future life, I’m quite afraid that if I will be really, really sick, antibiotics won’t work because I will be immune to them and that was the first choice for me. ”

Whilst a few discussed prioritization in terms of unpredictable impact on their careers and personal lives. For example, one participant talked about the effect of unexpected episodes on their career: “I lost my previous job because I was sick all the time and I was taking days off to go to see the GP or go to hospital because I had quinsy and that was awful.” Some discussed loss of opportunity due to the critical timing of the condition. However, tonsillitis also seems to have bothered others in a more social context, especially in terms of difficulty planning: “I had the same problem with that. I had plans and usually in the most important day I had tonsillitis. I was waking up and yes, my tonsils were swollen and I can’t speak, drink, eat, whatever. After a while when I was planning something I was thinking what if I will get tonsillitis? It was a nightmare.”

Theme 2 Trade off

A few participants used the strategy of trading one factor off against another in order to complete the ranking exercise. Participants discussed the trade-off in the context of balancing knowns and unknowns:

"For me, this seems to be all important because, on the piece of paper, you have to start from one and list them in a certain order, I think they're all important because even if certain things are not, how do you say, they are rare, you never know if you're going to be that case, so I think they are all quite important and things."

Theme 3 Reprioritization of less important factors

Another strategy that participants employed during the task was deprioritizing less important factors. Whilst some deprioritized symptoms that they had not personally experienced, others deprioritized based on factors they felt were less important.

When discussing how the lowest ranked items were chosen several participants talked about symptoms they had not personally experienced: "Well, last is definitely the sleep problems…" or another participant who said: "I think the sleeping problems, I didn't find it very…because most of the time my sleep was well."

Some participants deprioritized factors they did not feel were important. This participant when talking about why she placed cost as the lowest ranking said: "However much it costs every time wouldn't be that much of a reason to stop me from…that much of a reason for it to be more important, that wouldn't bother me but I can see why it could be a reason for other people." Or this participant when talking about why she placed immune function at the bottom of her list: "I mean, it's interesting but if it (the factor) doesn't have any impact then you just kind of ignore that point because you'll be thinking about something that does."

Theme 4 Clustering of factors based on perceived similarities

Participants also ranked factors they felt were similar together. The commonest grouping strategy was linking factors that were similar semantically or conceptually such as reducing episodes of sore throat and days of sore throat: "My second choice was, it's all under the same sort of bracket thing because my second choice was reducing unexpected episodes of a sore throat." Other participants had grouped through association and definition. One

participant felt that altering immune system was integral to the quality of life and grouped them together: “I’d say improving the overall quality of life and reducing the chance of altering the immune system because that’s essentially what it kind of does.” Others had used to the same reasoning to group quality of life with sleep, reduced taste with bad breath and costs with GP visits.

Did you have any difficulty with the instructions or terms used to describe the factors in the ranking exercise undertaken?

Most participants completed the prioritisation task with ease and within the 5 minutes. However, when probed about how the exercise could have been made easier they felt a more visual layout would have helped: “but because it was quite vague to me, I kind of needed it to be a bit more outlined and a little bit more, I don’t know, it kind of just felt like there could have been lots of different columns. It could have been, oh, mark one to ten before treatment, during or after.”

How well do participants understand terms used to describe evidence based factors relevant to the treatment of recurring sore throat?

When asked more specifically, about whether certain terms were difficult to understand, participants discussed three terms repeatedly:

1. Taste disturbance
2. Financial cost to community
3. Immunity

Taste disturbance

Half the participants had difficulty with the term taste disturbance: “Yes, the taste disturbance, again I just knew that it was something I hadn’t thought of so I didn’t need a clarification of what it was to know how important it was to me, but that was the only one I thought, well, I think I kind of was relying on thinking I understood it, not actually maybe understanding it”. When asked specifically about it participants “Oh, to taste something that you don’t necessarily want to…is it like that you don’t necessarily want to taste in your mouth as in just a foul taste? Is that what it means?”

Financial cost to the community

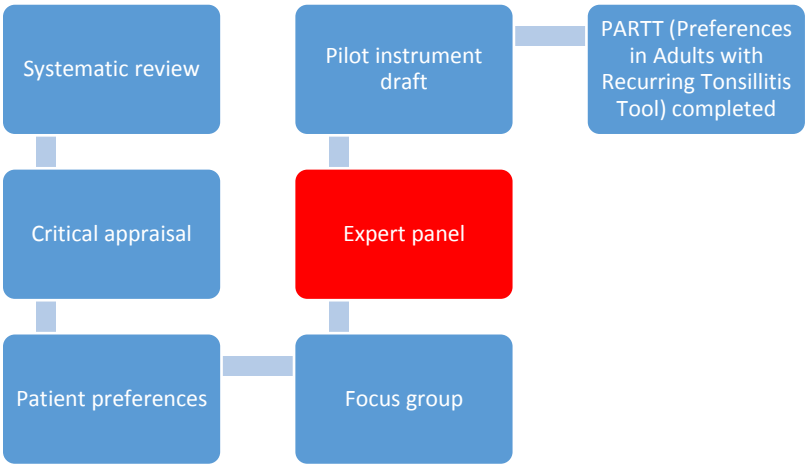
Whilst participants got the basic premise of this term they did not fully understand the full implications of the term: “Well, every time you had tonsillitis you had antibiotics right and you went to see the GP? Then the financial cost to the community is how much tonsillitis costs to treat.” And whilst this statement is true it fails to acknowledge that there is also a cost of the tonsillectomy to society and the overall cost can only be calculated when you know the difference between these figures.

Immunity

The original term was meant to refer to the idea that the tonsils are an important source of immune protection from lower respiratory and gastrointestinal infections. Therefore, following tonsillectomy you are at potential increased risk having have your immune protection affected which may result in respiratory infections or gastroenteritis. However, when participants were asked about this term they discussed immunity in terms of tonsillitis and not following a tonsillectomy: “As in like I had it a lot, so whenever I had tonsillitis, I have a rubbish immune system anyway, but it would be worse, like I’d be so ill and I’d easily get any other colds or anything, so how I understand it is just after the tonsillectomy, well, it has been for me anyway, when I, because I don’t have it anymore, I haven’t been getting ill as often, so it has reduced my immune system weakening, I don’t know if that was the thing…”

All other terms were understood in their correct context.

Expert panel: Shortlisting outcomes for use in the AHP



With regards to the treatment of adults with recurring tonsillitis I now had eleven outcomes for which I had 1. A defined effect size 2. An assessment of how credible and applicable the evidence based was for our patients, and 3. An assessment of how important they were to our patients. However, AHP requires that I have no more than seven outcomes to compare between to prevent fatigue(291). In addition, the outcomes should be easily understood by the respondent and not so conceptually like each that they prevent respondent discrimination.

Aim

Our aim was to reduce 11 treatment outcomes to seven or less based on patient importance, strength of evidence and effect size, dissimilarity and ease of understanding.

Method

An expert group of two ENT surgeons, one public health doctor, an AHP expert and psychometrician were presented the combined data from the critical appraisal and patient ranking exercises in a table (see Appendix Y – Tonsillectomy Outcomes Evidence and Patient Ranking)

The panel were given the brief that eleven outcomes needed to be reduced to seven or less. Criteria for removing outcomes included

- Insufficient evidence base or poor applicability to our population of interest
- Small effect of treatment on outcome
- Low patient priority
- Conceptually too similar to another outcome for patients to differentiate
- Difficult concept for patients to understand

I conducted a structured panel discussion by presenting findings related to 4 key topics and asking the panel to discuss with the overall aim of reducing 11 outcomes to seven or less. I structured my presentations to help the panel decide which outcomes were 1. more important to patients, 2. too similar to each other, 3. Too difficult to understand and 4. Most believable given the available evidence.

The expert panel were presented the results of the thematic analysis to provide insight into what themes were important for patients in the management of their recurring tonsillitis and how they evaluated the outcomes. For example, reducing days of sore throat, reducing episodes of sore throat, reducing days off-work and improving quality of life all mapped to three themes patient decision making: negative impacts of the condition, hope for a positive outcome, and individual trade off. These structured discussions helped the panel get a sense of which outcomes were more important to patients.

I presented outcomes that patients felt were clustered together. This helped the panel consider which outcomes were considered too similar to each other for valuable comparison, for example most patients felt that reducing days of sore throat and episodes of sore throat were the same thing.

I next presented outcomes that the patients had difficulty understanding, to assess whether they could be reworded or removed. The above structured discussions allowed the panel to appreciate the patients' perspective with regards to managing their recurring tonsillitis.

Finally, I presented the 11 studies reporting on their risk of bias (strong, moderate. weak) and effect size (small, medium, large) to help the panel decide how important tonsillectomy was

in relation to these outcomes (e.g. The panel felt that the effect size of tonsillectomy on voice was small and the credibility of the study weak).

Results

Final outcomes to be ranked by patients to determine their treatment preference
Reducing days of a sore throat
Reducing risk of haemorrhage
Halitosis reduction
Improvement in quality of life
Reducing visits to the GP

Table 25 Tonsillectomy outcomes chosen for inclusion in PARTT

This table shows the final five tonsillectomy outcomes chosen for inclusion in PARTT as selected by the expert panel, having removed six outcomes from the original list.

Overall 6 outcomes were removed by the expert panel based on conceptual similarity, poor evidence credibility or usability, small effect of treatment or low patient priority. The process is summarised below.

Two outcomes (days with sore throat and episodes of a sore throat) were considered too similar to each other conceptually based on focus group analysis (Theme 4 – Clustering of factors) and so a choice between excluding one or grouping both had to be made. Although they were both important, as evidenced by high patient rankings, the panel decided to use only one of these outcomes: reducing episodes of a sore throat, as it was ranked higher by patients.

Patients consistently ranked improving voice (11th), societal cost (10th) and immunological profile (9th) as the least important outcomes for them. The panel felt the evidence for these outcomes was poor and the effect size non-significant and so these three variables were excluded. In addition, taste disturbance (6th) and snoring (8th) were also ranked low by

patients and the panel felt the evidence for these outcomes was insufficient and so they were excluded.

This left five patient-chosen outcomes that have sufficiently strong evidence: Reducing episodes of a sore throat, reducing visits to the GP, improving the quality of life, reducing halitosis and reducing haemorrhage risk – see Table 25.

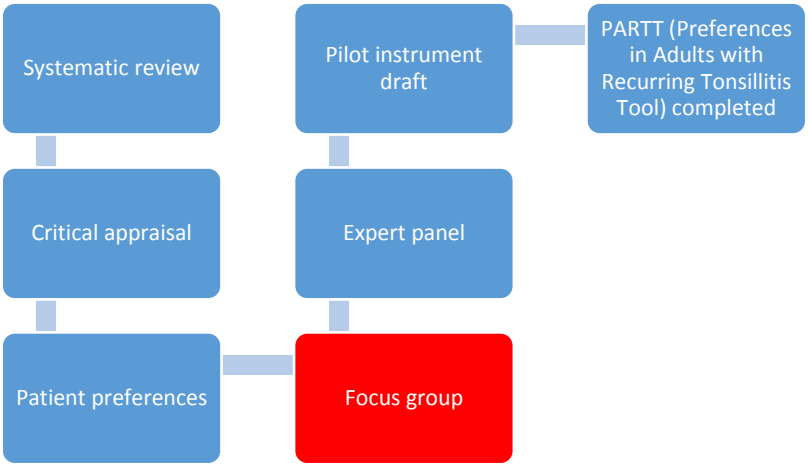
Summary

I undertook a systematic review and critical appraisal of randomised and nonrandomised studies investigating treatments for recurring tonsillitis in adults to shortlist outcomes that were important to patients and had strong evidence. And whilst there were thousands of studies investigating outcomes from treatments for recurring tonsillitis, the majority related to intra-operative or early postoperative outcomes that I felt were more important to the surgeon and anaesthetist than to patients making their decision.

The evidence of all eleven outcomes from the systematic review and critical appraisal, patient rankings and thematic analysis of focus group was discussed with an expert panel. The expert panel used the above information to select five outcomes for which there was at least moderate supporting evidence and patients also felt were important.

The following section describes how I converted the patient important five treatment outcomes into an AHP based questionnaire, designed to elicit treatment preference for adults with recurring tonsillitis.

Drafting and piloting the Preferences in Adults with Recurring Tonsillitis Tool (PARTT)



Aim

To design and pilot an AHP based instrument that can be used to elicit treatment preferences in adults with recurring tonsillitis.

Methods

I designed the Preferences in Adults with Recurring Tonsillitis Tool (PARTT) based on standard AHP methodology that asks respondents to undertake pairwise comparisons between outcomes, based on their personal preferences. The method required all five outcomes be compared to each other on a response scale with semantic response categories (Equal preference between two outcomes, Moderate preference, Strong preference, Very strong preference and Extreme preference), resulting in 15 pairwise comparisons in total. In the introduction, a short explanatory sentence was provided for each outcome so that the respondent could weigh up the potential size of the benefit or risk when making their preference choice. In addition, I provided instructions of how the exercise should be undertaken. Once the PARTT was designed I piloted it with adults with recurring tonsillitis at the lead author's clinic.

I approached all patients attending a weekly sore throat clinic (which I had organised and ran), between 01/11/2013 and 01/12/2013, to help refine the instrument. The study was verbally

explained to patients and consent was verbally obtained. Patients were asked to use the think aloud technique as they read and completed the questionnaire(311).

I took notes whilst patients completed the questionnaire and made appropriate changes to the instrument based on observations and represented the revised instrument to the next batch of patients. The process was repeated until the patients seemed to easily understand the instructions and the outcomes that needed to be compared.

Results

The draft PARTT is presented in Appendix K - Draft version of PARTT.

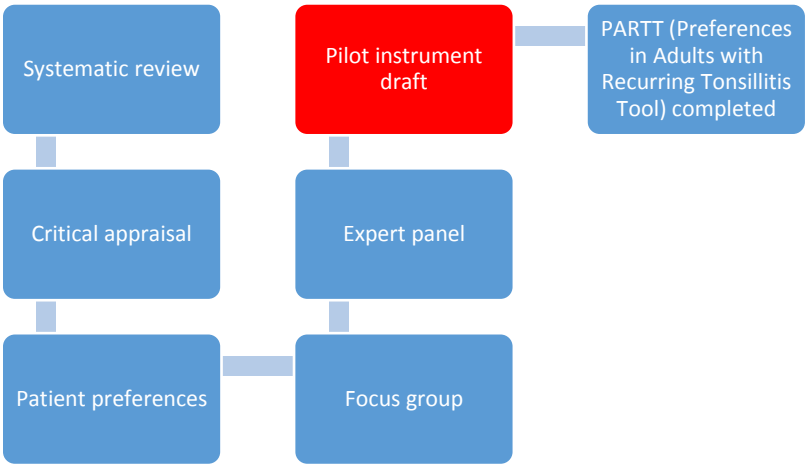
Twelve patients, who had recurring tonsillitis, were approached to take part in the pilot in the sore throat clinic by the lead author. All patients agreed to participate and provided verbal consent.

The first three patients had considerable difficulty understanding the instructions. Therefore, the lead author verbally explained them so that they could proceed onto the comparisons. One patient also found the font on the comparison scale too small to read. An example was added into the instructions and the font size was increased.

The following four patients had no problems with the instructions, although one complained that the instruction took him a long time to read, and he lost interest several times. All four patients found it tiresome to have to keep returning to the introductory paragraph to reference the likely chance of risk or benefit in relation to each outcome.

Therefore, I changed the example to something more abstract but interesting (i.e. trying to find the best way to go to the shops). Additionally, I also added the explanatory sentence of the chance of risk or benefit related to each outcome under every comparison.

The resulting instrument was tested on five patients over two weeks with no changes made to it. All patients understood the instructions, felt the example helped, and undertook the comparisons with ease. The final version of the PARTT is shown in Appendix L – Final PARTT.



PARTT description

As described previously the PARTT requires study participants to undertake 15 pairwise comparisons between 5 potential treatment outcomes, based on their personal priorities. Verbal scale responses are converted to the Saaty scale so that equal preference between two outcomes would score one, Moderate preference- 3, Strong preference– 5, Very strong preference- 7 and Extreme preference- 9.

Scores derived from the 15 comparisons are used to calculate preference ranks for each of the five potential treatment outcomes using standard AHP techniques(304) (described in Appendix T AHP methodology and Ranking results). Higher ranks indicate higher preference, hence outcomes ranked number 5 are the most important and those ranked 1 are the least important to respondents- these are called the Ranked Outcome Priorities(R.O.P).

The consistency of participants' pairwise comparisons is also determined, allowing those who had answered questions inconsistently to be removed from the cohort. Consistency scores are calculated through standard AHP algorithm (274).

Finally, a tonsillectomy preference score can be calculated based on standard AHP methodology. This is a sum of the participants' preferences for each outcome, weighted to reflect that likelihood of outcome occurrence if tonsillectomy was chosen. For example, if a participant has a high preference for reducing halitosis but the likelihood that tonsillectomy would reduce halitosis was low, based on the literature, then the impact of halitosis reduction on the tonsillectomy preference score would be down-weighted (see Appendix T AHP

methodology and Ranking results). Participants' tonsillectomy preference score is calculated on a continuous scale between 0-1, with higher numbers indicating greater preference. The participant's 'conservative treatment' preference score is the converse of the tonsillectomy preference score, so the greater the preference for one treatment the lower the preference for the other. If a participant has given all outcomes equivocal priority than the tonsillectomy preference score would be 0.50.

Consistency checking

Participants scoring greater than 0.5 on the consistency ratio are excluded from the analysis as their comparisons are deemed random. Sensitivity analyses were conducted for subgroups if the cohort was restricted by consistency ratios less than 0.2, 0.3, 0.4 and the whole sample.

Summary

I designed the PARTT based on the Analytical Hierarchical Process – a method that is based on rational decision making, but allows for and quantifies inconsistent decision making that is common in real time. The method allows a valid method to measure preferences that is easy to deploy and has low cognitive burden.

I systematically searched and critically evaluated the evidence base to ensure I had detailed information on treatment outcomes – in terms of how they are affected based on the treatment chosen and how credible and usable the evidence was in relation to our patients of interest. The eleven outcomes identified were presented to patients with recurring tonsillitis and they were asked to rank them in terms of their importance to their treatment decisions. Additionally, I undertook a focus group to better understand how patients were making their treatment decisions. The above information was presented to an expert panel with the goal of reducing eleven outcomes to seven or less. The panel removed outcomes based on patients' priorities, patients' conceptually finding outcomes too similar to each other, poor quality evidence or small effect of treatment on outcomes. Five outcomes were identified to use in an instrument that uses AHP to elicit treatment preferences in adults with recurring tonsillitis.

A pilot instrument – PARTT - was designed and sequentially tested on patients with recurring tonsillitis and changes made incrementally until it was considered to easily and efficiently elicit treatment preferences.

The following chapter will describe our observational study of the role of treatment preference (using PARTT) and decisional uncertainty (using the DCS, described in Chapter 5) on decision-making in adults with recurring tonsillitis.

The following chapter will describe our observational study of the role of treatment preference (using PARTT) and decisional uncertainty (using the DCS, described in Chapter 5) on decision-making in adults with recurring tonsillitis.

CHAPTER SEVEN: A cohort study to investigate the roles of decisional uncertainty and preferences in explaining regional tonsillectomy rate variation

Chapter outline

In this chapter I have presented an observational study investigating treatment decisions for adults with recurring tonsillitis. The study investigated the role of patient and surgeons' decisional uncertainty using the Decisional Conflict Scale - DCS) and treatment preference (using Preferences in Adults with Recurring Tonsillitis Tool - PARTT) on the management chosen for adults with recurring tonsillitis. Overall decisional uncertainty was low for patients and ENT surgeons. However, decisional uncertainty was significantly higher for both patients and ENT surgeons involved in consultations that resulted in watchful waiting compared with those that ended in which tonsillectomy was chosen. Patient PARTT scores were not associated with treatment chosen. However, surgeons' PARTT scores were related to treatment chosen, even after controlling for patient preference and disease severity.

Introduction

I have shown that the rates of sore throat in the community and in primary care, along with tonsillectomy rates in secondary care change with factors that are known to vary between regions (specifically age, sex, ethnicity, presence of chronic medical diseases, number of people in the household, smoking status and population density). Controlling for these population factors reduced regional tonsillectomy, recurring sore throat and community sore throat variations. Following critical review of the literature (17,22,23,39-41,271,312) and my subsequent systematic review (100) around this topic, I have hypothesised that factors during the medical consultation may also have a part to play in variations observed.

In relation to recurring tonsillitis and tonsillectomy there is only very preliminary evidence from observational studies (17,23) that investigates the role of the consultation on surgical rate variation. In general, in relation to other surgical conditions, such as breast cancer, the literature suggests that there are two potential drivers related to the medical consultation that plays a role in regional surgical rate variation: professional uncertainty and treatment preference.

Professional uncertainty, decisional uncertainty and Decisional Conflict Scale (DCS)

Observational studies investigating the rates of surgery have shown greater variation for procedures which lack professional consensus on how cases should be managed (24-26), with tonsillectomy being one of them(17). Lack of professional consensus has been described as professional uncertainty, which has never before been conceptually defined, rather inferred from general lack of professional consensus. Additionally, it has never been measured directly only inferred through survey studies that show large variation in the manner doctors would manage hypothetical patients (see chapter 1). Currently, there is no study that has investigated how professional uncertainty in the management of recurring tonsillitis manifests itself during the shared decision making process between ENT surgeon and patient to align decisions towards tonsillectomy in some regions more than others.

It has been hypothesised that in the absence of strong professional consensus surgeons may be more uncertain on what to advise and patients may feel more uncertain about what treatment to choose (161). I have hypothesised that lack of professional consensus manifests

itself during shared decision making, between ENT surgeon and patient, specifically as decisional uncertainty, for both surgeons and patients. Decisional uncertainty is a well-defined construct within shared decision-making paradigm and can be measured using the Decisional Conflict Scale (DCS) (263) for patients, or Provider's Decision Process Assessment Instrument (PDPAI)(269) for surgeons. The DCS, and surgeon specific DCS called PDPAI, have strong psychometric properties with regards to validity and reliability.

Surgeon treatment preference, patient treatment preference and Preference in Adults with Recurring Tonsillitis Tool (PARTT)

Based on a seminal cross sectional study of paediatric tonsillectomy rates by region (16)and qualitative study of recurring tonsillitis consultations by region (23), Wennberg (24) hypothesised that regional tonsillectomy rate variation resulted from regionally aligned surgeon preference, a theory he coined the "Surgical Signature". If patients who see surgeon A are more likely to receive tonsillectomy than if they saw surgeon B, it can be hypothesised that surgeon A has a treatment preference for tonsillectomy. However, in the modern health care setting of shared decision making, where the surgeon incorporates the patients' preferences, with the best available evidence, to help reach a treatment decision rather than paternalistically dictating the treatment, surgeon preference is a complicated construct. Consultations with surgeon A could have resulted in tonsillectomy being chosen more frequently because the surgeon perceived tonsillectomy was a more effective treatment for certain outcomes, perceived patients prioritised these outcomes more than others, proxy bias, or the surgeon's personal preference towards tonsillectomy. Partly due to the complexity of this construct surgeon treatment preferences, as they apply in the medical consultation, have always been difficult to measure on a scale large enough to inform discussions about regional surgical rate variations. Instead they have been implied from smaller qualitative studies, especially around the surgical management of breast cancer (26). There is no evidence that directly reports on surgeons' treatment preferences for the management of adults with recurring tonsillitis, and how this affects treatment decisions.

In relation to regional surgical rate variation, whilst surgeon treatment preference has a historical context and has been discussed (disproportionately in relation to the level of current evidence) patient treatment preferences have only recently come to the fore. There

is now a growing body of evidence from studies of orthopaedic conditions that regional surgical rate variations are associated with regionally aligned patient treatment preferences (78). (78,81-83) (85-88), (272) (90) (91). However, due to the difficulties in eliciting patient treatment preferences efficiently on a large scale, preferences have been implied by willingness for treatment based on selecting an ordinal value on a spectrum from "*definitely not willing to have to treatment*", through to "*definitely willing to have treatment"*, has never been tested with respect to validity or reliability. There is no evidence that directly reports on treatment preferences for adults with recurring tonsillitis, and how these patient preferences influence the treatment decision.

Recent advances in decision theory have led to rapid progress in the elicitation of preferences, and the analytical hierarchical process (AHP) offers a rapid and reliable method to elicit both patient and surgeon treatment preferences. Using outcomes from treatment that are important to adults with recurring tonsillitis I designed an instrument based on AHP that is capable of eliciting patient treatment preferences – PARTT. By asking a surgeon to complete PARTT from a typical recurring tonsillitis patients' perspective it is possible to approximate the surgeon's treatment preference in the context of the medical consultation, as s/he must balance their beliefs around the effectiveness of tonsillectomy with the outcomes they perceive are important to their patients. In this way, I can quantify surgeon treatment preference, in the context of a medical consultation, without reducing its complexity to 0-10 visual analogue scale asking surgeons how much they like tonsillectomy.

Aims

To investigate the role of decisional uncertainty and treatment preference on the treatment chosen for adults with recurring tonsillitis

Objective

1. To define the relationship between decisional uncertainty and treatment selected for recurring tonsillitis from the perspective of both the patients and ENT surgeons

2. To define the relationship between ENT surgeons' decisional uncertainty and the patients' decisional uncertainty

3. To define the relationship of post-consultation patients' treatment preference and treatment chosen

4. To define the relationship between pre-consultation surgeons' treatment preference (in the context of a medical consultation) and the treatment chosen by the patient

Method

Ethics

Ethical approval for this study was granted by Nottingham Research Ethics Committee (15/EM/0191) on 8th May 2015.

Sample

Hospital selection

Inclusion criteria related to whether a hospital had regular ENT clinics and sufficient volume of recurring tonsillitis patients to recruit patients. There was a local and national approach to recruiting sites. The national approach was to advertise the study on the NIHR research portfolio and send an email out to CRN ENT leads. Locally, I recruited sites by advertising the study amongst the regional trainee research collaborative – Otolaryngology Trainee Investigators Collaborative (OTIC). OTIC has 46 trainee surgeon members across 14 hospitals in the Greater London region. A total of 17 sites (5 nationally and 12 locally) demonstrated interest in participating in the study. They were provided with study protocols, minimum recruitment targets per site and asked to discuss recruitment with their Trust's Research and Development departments to assess feasibility. Sites that felt they had the capacity to recruit the minimum number of patients were asked to contact me.

Surgeon selection

The Principal Investigator (P.I.) was appointed as the Consultant who had shown an interest in participating in the study. I briefed the P.I. regarding study etiquette and protocol, personally at each site, before recruitment started. The P.I. at each site was asked to discuss the study with all ENT surgeons in the department, to this end, each surgeon was given an

information sheet about the study and asked to read it (see Appendix M - Detailed Doctor Information Sheet). The P.I. explored his/her colleagues' willingness to participate in the study. I sent those who were interested a consent form to complete (see Appendix N - Doctor Consent Form). The P.I. recruited a local research team that included junior doctors, research trained nurses (depending on local availability) to help conduct the study.

Patient selection

Consecutive patients attending ENT clinics were recruited if they were 16 years old and over with recurring tonsillitis episodes frequent enough to justify a tonsillectomy (as judged by the participating ENT surgeon according to SIGN guidelines). Patients referred for tonsillectomy for snoring or halitosis, or with insufficient English language skills to undertake the consent process were excluded. The local research team co-ordinated screening of GP referral letters to identify potential participants. Potential participants were sent information by the clinical team about the study prior to their hospital appointment (see Appendix O - Patient Invitation Letter). On the day of their appointment if they were found to meet inclusion criteria, as ascertained by their consulting surgeon, they were asked if they were happy to discuss study participation with a member of the local research team (defined above). A member of the local research team discussed the study in more detail and provided a detailed information sheet (see Appendix P – Detailed Patient Information Sheet). Patients who were happy to participate were asked to complete a consent form (Appendix Q – Patient Consent Form).

Sample size calculation

The study was powered to detect a clinically significant decisional conflict score of 25/100 (or an effect size of 0.4).This figure is based on initial validation studies in e adults who were offered influenza vaccination showing that the DCS can be used to detect an effect size of 0.4 (or a DCS score of greater than 25 out of 100) between those who postpone their decision and those who actively undertake treatment (313). A subsequent meta-analysis of ten randomised controlled trials showed that an effect size of 0.4 (or a total score of greater than 25 out of 100) in the DCS can differentiate between those who delayed their decisions and those that did not, and those who regretted their decision and those that did not(266). A prospective study of 100 aneurysm patients facing a treatment decision showed that an effect size of 0.4 could predict behaviours of decision postponing(267). I therefore powered our

study to detect a moderate effect (effect size=0.4) based on these studies. I predicted that less than one third of patients would choose conservative therapy (3) and estimated the standard deviation of the DCS in our population conservatively at 0.8 (314,315). Therefore, using two-sided significance level of 0.05, a power of 0.8 I estimated that that I would need to recruit 150 patients (100 who chose tonsillectomy and 50 who chose conservative therapy).

Instruments

Decisional uncertainty (DCS&PDPAI)

The DCS and the surgeon adapted DCS (PDPAI) both have 16 items scored on a five-point scale (0=Strongly agree, 4= strongly disagree). The total score, and sub scores, are transformed to 0-100 scale, with higher scores associated with greater uncertainty. The scale discriminates between those who make and those who delay decisions (effect sizes range from 0.4 to 0.8). The instrument has a minimal clinically important difference previously established; patients scoring 25 or more are likely to delay decisions, whereas those with a score of less than 25 tended to make decisions (264). Patients scoring more than 25 out of 100 (265) on the total patient DCS has been related to decisional delay, departure from active treatment, decision regret, nervousness and a higher intention to sue physicians in cases of harms from treatment and has been the score most commonly used to distinguish a harmless from a harmful level of decisional conflict (265-268). Decisional uncertainty scores for both patients and surgeons were logarithmically transformed to approximate a normal distribution. Parametric analyses were conducted on the transformed variables.

Preference in Adults with Recurring Tonsillitis Tool (PARTT)

PARTT requires respondents to make a 15 pairwise preference comparisons between 5 potential treatment outcomes (reducing days of sore throat, reducing risk of haemorrhage, halitosis reduction, improvement in quality of life, and reducing visits to the GP) on a verbal response scale (Equal preference, Moderate preference, Strong Preference, Very Strong Preference, Extreme Preference). There are four final outputs (process described Appendix T AHP methodology and Ranking results):

1. Ranked outcome priorities: Potential treatment outcomes ranked 1-5 in decreasing order of preference (e.g. reducing days of sore throat=1, reducing visits to GP=5 suggests respondent's main priority was reducing days of sore throat and lowest priority was reducing visits to the GP)

2. Outcome preference score: Each potential treatment outcome scored between 0-1 to demonstrate magnitude of preference between them (e.g. reducing days of sore throat=0.4, reducing visits to GP=0.1 suggests that reducing days of sore throat was 4 times more important to respondent than reducing visits to the G.P)

3. Consistency score: Overall score describing how consistently comparisons were completed, scored between 0-1, with scores greater than 0.2 suggesting inconsistent comparisons and greater than 0.5 suggesting randomly undertaken comparison

4. Tonsillectomy preference score: Overall score, between 0-1, describing how respondent's values align with tonsillectomy (e.g. respondent A=0.2, respondent B=0.5, respondent C=0.7 suggests that respondent A has values that do not align tonsillectomy, respondent B has values that do not indicate an obvious tonsillectomy preference or aversion, respondent C has values that align more closely with tonsillectomy).

Whilst all four PARTT scores were calculated and analysed with relationship to treatment decision chosen only the consistency and tonsillectomy preference scores are presented below as they demonstrate key findings. For results relating to the ranked outcome priorities see Appendix P.

Data Collection

Surgeon data

Consenting ENT surgeons were given a unique surgeon identifier code by the local research team. Participating ENT surgeons completed a one-time questionnaire booklet that included the PARTT, under the hypothetical assumption they were a patient with recurring tonsillitis and basic demographic questions (see Appendix R – Doctor pre-consultation questionnaire). Immediately, following on from every consultation the surgeon undertook with a study patient they were asked to complete the PDPAI (DCS for surgeons) (Appendix S – Doctor post

consultation questionnaire). Both questionnaires had the ENT surgeon's unique identifier code attached so their scores could be compared later.

Patient data

Patients who agreed to participate were given a patient unique identifier and asked to complete a questionnaire booklet immediately following on from their ENT consultation, by the local research team. The booklet included the DCS, PARTT, satisfaction and basic demographic questions. The patient unique identifier and the surgeon unique identifier (of the surgeon they had seen) was added to the booklet so that scores could be compared later.

Analysis

Sample

Chi squared testing was used to compare available characteristics of our surgeon participants with ENT UK surgeon membership and patient participants with both CALIBER and National Prospective Tonsillectomy Audit participants.

For analyses of patient and surgeon treatment preferences, participants scoring greater than 0.5 on the PARTT consistency score were excluded from the analysis as their comparisons were deemed random. Sensitivity analyses were conducted for subgroups if the cohort was restricted by PARTT consistency scores of less than 0.2, 0.3, 0.4 and the whole sample.

For analyses regarding decisional uncertainty, participants and resulting consultations were excluded if either the surgeon or the patient had missing data with regards to the PDPAI or DCS, respectively.

General results

Patient and surgeon sociodemographic as well as hospital characteristics were investigated in relation to the treatment chosen using Chi Squared tests.

The association between patient and surgeon decisional uncertainty and treatment chosen

To address the first objective, I developed univariable logistic models that investigated the role of the logarithmically transformed decisional uncertainty scores on the likelihood of consultation resulting in a tonsillectomy being chosen. I developed multilevel multivariable mixed effect models that measured the role of logarithmically transformed uncertainty scores

on the treatment chosen, after controlling for disease, patient, surgeon and hospital factors that I had considered to be conceptually related to the outcome. I forced consultations to be nested within ENT surgeons and ENT surgeons to be clustered within the hospital, to account for random variation at these levels. ENT surgeon and hospital variables were added sequentially and hierarchically, as random effects to the appropriate level, if they improved the model fit (as assessed by Akaike's information criterion and pseudo-R squared values). If they did not improve upon the model they were added as fixed effects. I developed multi-level mixed effects linear regression models to assess the role of patient, ENT surgeon and hospital variables on the change in patient and then surgeon uncertainty scores (logarithmically transformed). Models were developed using the same methodology described above.

The association between ENT surgeon's decisional uncertainty and patient's decisional uncertainty

To answer the second objective, I undertook tests of correlation between the patient (DCS) and surgeon (PDPAI) uncertainty scores. I used Pearson's correlation coefficient to describe the correlation between patient and surgeon uncertainty scores.

The association between tonsillectomy preference score and treatment chosen

I undertook cluster analyses to investigate the role of tonsillectomy preference scores on treatment chosen (tonsillectomy or watchful waiting), for surgeons and patients separately. The goal of clustering was to assign the respondents into clusters which are grouped with similar characteristics (316). Clustering analysis has already been used in many application domains such as market research (317), however, the popularity has recently grown in health care preferences (318). When clustering, one is interested in grouping respondents that cannot be distinguished from each other, and separating those that can(319).

I chose to undertake cluster analysis as there are no known classifications of tonsillectomy preference scores (i.e. we do not know which score identifies high preference and which low preference patients, or even if there is moderate preference band). The clustering method chosen for this study was Ward's or minimal increase of sum-of-squares, which used squared Euclidean distance between data points. This is the clustering method most commonly used to create a classification across people. I used standard criteria to check the validity of my

clustering methodology (internal validity: Calinski-Harabasz, Silhouette and Duda criteria; external validity: True treatment chosen; cross validity: compare clusters grouping created by different cluster techniques e.g. median partitioning).

I undertook multi-variable logistic analyses to assess the relationship of participant cluster groupings, for tonsillectomy preference score with actual treatment chosen, for patients and ENT surgeons separately. Variables that were associated with treatment choice were added sequentially and hierarchically to the model to the patient or surgeon specific models.

Finally, I added both patient and ENT surgeon preference clusters into the multivariable model to investigate the role of both ENT surgeon and patients' treatment preference score on the treatment chosen.

Results

Sample

Hospital selection

Eight sites in the London area and six nationally agreed to participate in the study. Geographically the study included sites from Greater Manchester, Oxfordshire, Norwich and Norwich, Hertfordshire, Buckinghamshire, Poole and Greater London. Four (29%) sites were teaching hospitals (national proportion 3%) and ten district general hospitals.

Surgeon selection

All ENT surgeons of recruiting hospitals, consented to participate. Five surgeons did not complete the post consultation questionnaires. Therefore, five consultations lacked surgeons' uncertainty score and were excluded from uncertainty analyses. Ten surgeons had inconsistent PARTT responses, and so were excluded from the study (consistency ratio>0.5). Therefore, I analysed preferences results from 160 consultations. Two thirds of ENT surgeons had a consistency ratio of less than 0.2 and 90% had a ratio of less than 0.4. Sensitivity analyses showed no association between outcome rankings and treatment chosen, irrespective of consistency ratio cut off threshold.

Surgeon characteristics are shown in the Table 26. Age, sex and grade of the participating ENT surgeons were comparable to that of the national ENTUK membership (t-test p>0.1).

However, there were no data available for ENTUK member ethnicity, years of practice or type of hospital worked in.

ENT surgeons' variables	Percentage in this subgroup (n)
Hospital type	
University	42%(26)
District General Hospital	58%(36)
Age categories	
20-29	8%(5)
30-39	31%(19)
40-49	35%(22)
>50	21%(13)
Missing	5%(3)
Gender	
Male	64%(39)
Female	31%(19)
Missing	3%(2)
Ethnicity	
White	47%(29)
Non-white	47%(29)
Missing	6%(4)
Grade	
Consultant	41%(25)
Registrar	28%(17)
Staff Grade	11%(7)
Associate Specialist	10%(6)
Core trainee	5%(3)
Missing	3%(2)
Years of ENT practice	
>1	3%(2)
1-5	15%(9)
6-10	21%(13)
11-20	36%(22)
>20	20%(12)
Missing	3%(2)

Table 26 ENT surgeons' demographics

This table shows the ENT surgeons' demographics. Age, sex and grade of the surgeons was comparable to the national average (n=60).

Patient selection

See Figure 19 Patient recruitment flow diagram.

Screening of GP letters identified 329 potential participants. Twenty-eight patients did not attend their appointments (n=301). One hundred and three patients were excluded as they did not meet inclusion criteria (n=198). Eight patients could not be recruited as the research team was unavailable on the date of their outpatient appointment (n=190). From the 190 patients that were eligible for inclusion six patients did not want to participate in the study (n=184). Since I successfully recruited 184 patients, our study had a recruitment rate of 98%. Only one participant had missing data in their DCS and so their results were excluded from the uncertainty analyses. Seven patients did not complete their PARTT completely, whilst 17 were inconsistent with their responses (consistency ratio>0.5) and so were excluded from preference analyses. Two thirds of patients had a consistency ratio of less than 0.2 and 90% had a ratio of less than 0.4. Sensitivity analyses showed no association between outcome rankings and treatment chosen, irrespective of consistency ratio cut off threshold. Therefore, the whole cohort was used for subsequent analyses (Table 29 Sensitivity Analysis for consistency ratios).

Recruited patients were not significantly different from those not recruited in terms of age (p=0.51) and sex (p=0.65) - tested using two-sample test of proportions. The distribution of patient characteristics across 184 recruited patients was compared to characteristics reported in the National Prospective Tonsillectomy Audit (NPTA) (56) (n=33,680) for gender and our CPRD-HES study of adult tonsillectomy (n=6830) for age and ethnicity (since this data was not available for the NPTA) using two sample test for proportions. Gender distribution in our study was comparable to NPTA (83% vs 85% female respectively p=0.45). Age and ethnicity distribution were comparable to those who had tonsillectomies recorded on our CPRD-HES database (72% vs 68% 16-29-year-old respectively, p=0.25; 77% vs 75% white ethnicity, respectively, p=0.54).

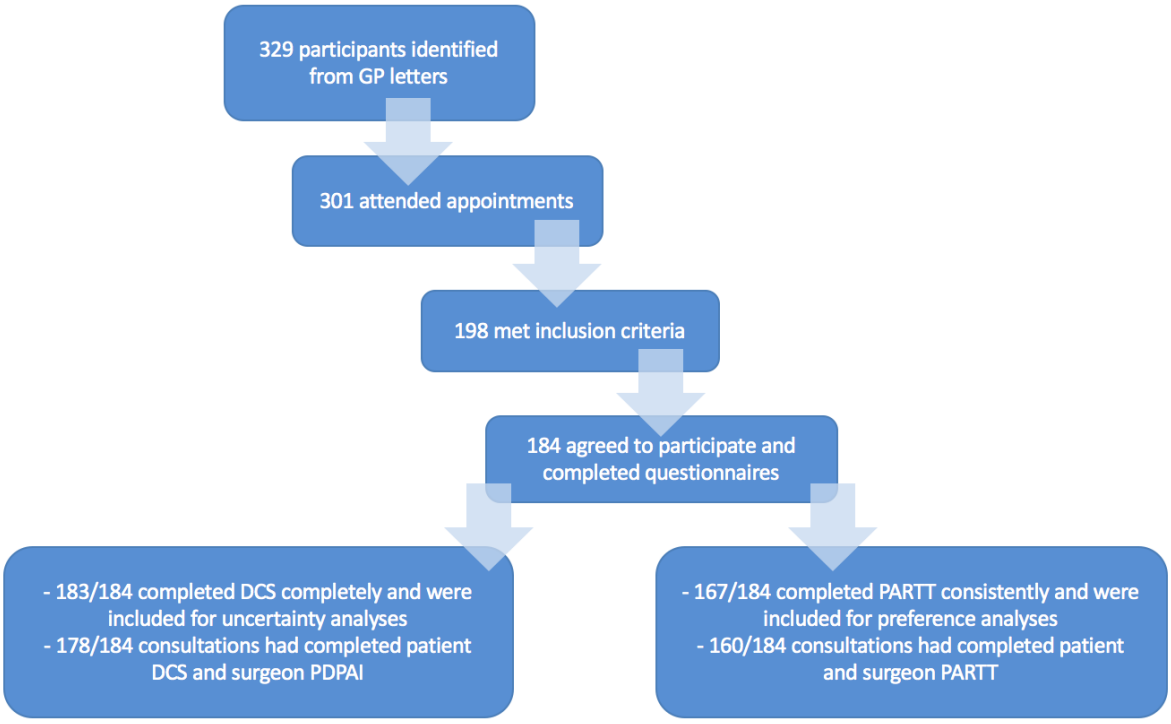


Figure 19 Patient recruitment flow diagram

This figure shows the number of patients who were initially identified from their GP letters, culminating in the number who went on to complete DCS and PARTT.

From the 184 patients recruited, 36 (20%) chose to treat their recurring tonsillitis non-surgically (watchful waiting whilst 148 (81%) chose to undergo a tonsillectomy. The distribution of patient characteristics across this treatment choice can be seen in Table 27 below. There was no statistical difference in patient and disease characteristics between the two groups.

Patient variables	Percentage who chose conservative (n)	Percentage who chose tonsillectomy (n)	Chi Squared P-value	Percentage of total
Hospital type				
University	22% (13)	78% (47)	0.67	33%
District General	19% (23)	81% (101)		67%
Age categories				
16-29	18% (24)	82% (109)	0.39	72%
30-59	24% (1)	76% (35)		25%
Missing	20% (1)	80% (4)		3%

Gender				
Male	16% (5)	84% (27)	0.55	17%
Female	20% (30)	79% (118)		80%
Missing	25% (1)	75% (3)		3%
Ethnicity				
White	17% (25)	82% (118)	0.19	77%
Non-white	27% (10)	73% (27)		20%
Missing	25% (1)	75% (3)		2%
Patient perceived appropriateness of GP referral timing				
Prompt	14% (7)	86% (44)	0.13	27%
Appropriate	25% (23)	74% (67)		49%
Delayed	13% (5)	87% (33)		20%
Missing	20% (1)	80% (4)		3%
No. of tonsillitis episodes in last 12 months				
0-5	22% (13)	78% (17)	0.83	12%
6-7	17% (10)	83% (49)		32%
>7	19% (12)	81% (50)		33%
Missing	20% (1)	80% (4)		3%
No. of days off in last 12 months				
0-5	25% (13)	75% (38)	0.52	28%
6-10	21% (10)	77% (37)		26%
>11	15% (12)	85% (67)		43%
Missing	14% (1)	86% (6)		4%
Total	20% (35)	80% (144)		100%

Table 27 Patient demographics by treatment chosen

This table summarises the demographics of the patient groups in both the tonsillectomy and conservative treatment group (n=184).

These patients were sent information sheets prior to their ENT consultation. On the day of their ENT appointment the local recruitment officer approached the ENT surgeon to alert them of a potential study participant. Twenty-eight patients did not attend their appointments. For patients who did attend eligibility of study inclusion was assessed during the consultation by the ENT surgeon. One hundred and three patients were excluded as they did not meet inclusion criteria. For those who met the inclusion criteria the ENT surgeon asked if they would be happy to talk to a member of the recruitment team. Eight patients could not

be recruited as the research team was unavailable on the date of their outpatient appointment. From the 198 patients that were eligible for inclusion six patients did not want to participate in the study. Since I successfully recruited 184 patients, our study had a recruitment rate of 98%.

Recruited patients were not significantly different from those not recruited in terms of age (p=0.51) and sex (p=0.65) - tested using two-sample test of proportions.

The distribution of patient characteristics across 184 recruited patients was compared to characteristics reported in the National Prospective Tonsillectomy Audit(NPTA) (56)(n=33,680) for gender and our CPRD-HES study of adult tonsillectomy (n=6830) for age and ethnicity (since this data was not available for the NPTA) using two sample test for proportions. Gender distribution was comparable to National Prospective tonsillectomy audit (83% female in our study compared to 85% in NPTA, p=0.45). Age distribution was comparable to tonsillectomies recorded on our CPRD-HES database (e.g. 68% in the CPRD-HES database compared to compared to 72% in our study sample were 16-29 year olds, p=0.25). The white to non-white ethnicity proportion of our sample was comparable to our CPRD-HES database (75% white in CPRD-HES and 77% white in our study, p=0.54).

General results

From the 184 patients recruited, 36 (20%) chose to treat their recurring tonsillitis with conservative therapy whilst 148 (81%) chose to undergo a tonsillectomy. The distribution of patient characteristics across this treatment choice can be seen in Table 27 Patient demographics by treatment chosen. There was no statistical difference between patient and disease characteristics on treatment chosen.

The association between patient and surgeon decisional uncertainty and treatment chosen

Median patient uncertainty was low at 12.2 (interquartile range 0-25), with a skewness score of +1.90. Median ENT surgeons' uncertainty scores were higher than patients at 36.67/100 (interquartile range 26.67-43.33), with a skewness score of +0.8. Since both patient and ENT surgeons' scores were positively skewed I logarithmically transformed the scores. Therefore, overall, surgeons were three times more uncertain than patients.

There was a moderate effect (Cohen's D=0.5) of patient decisional uncertainty between those who chose tonsillectomy and those who selected watchful waiting – higher uncertainty was associated with choosing watchful waiting. ENT surgeons showed a slightly larger effect for decisional uncertainty when consultations ended in watchful waiting being selected (Cohen's D=0.7).

Patients' (adjusted OR 0.41 95% CI 0.18-0.93) and ENT surgeons' (adjusted OR 0.0001 95% CI <0.001-0.08) uncertainty scores were independently lower when conservative therapy was chosen, even after controlling for disease factors (number of tonsillitis episodes and days off work in last 12 months), patient factors (age, gender, and ethnicity), ENT surgeon factors (surgeon age, sex, ethnicity, grade and years of practice) and hospital factors (type of hospital). Whilst the multi-level model showed clustering of results at the ENT surgeon level (variance 1.27 95%CI 1.56-10.39), there was no measurable clustering at the hospital level (variance <0.001).

The association between ENT surgeon's decisional uncertainty and patient's decisional uncertainty

There was no correlation between patient and surgeon uncertainty scores following tonsillitis consultations (Pearson's correlation coefficient=0.09).

The association between tonsillectomy preference score and treatment chosen

ENT surgeons reported a mean tonsillectomy preference score 0.57 (95%C.I 0.54-0.58) and patients with recurring tonsillitis reported a mean tonsillectomy preference score of 0.55 (95% CI 0.54-0.57). Student's T test did not show any statistical difference in patient tonsillectomy preference scores between these two treatment groups (P=0.48).

Whilst there was no difference in the mean tonsillectomy preference scores between ENT surgeons and patients, there was also no correlation between individual preference scores of patients and their consulting ENT surgeons: Pearson's Correlation coefficient 0.14(95% CI: -0.07-0.35).

Patients' tonsillectomy preference scores are displayed in Figure 20 Patients' tonsillectomy preference scores. Calinski pseudo F score index (237.93), Duda's Je(2)/Je(1) index (0.71) and Silhouette scores (0.69)-were all highest for 3 clusters, suggesting strong internal validity of

these clusters. see Figure 21. Table 30 displays the characteristics of each cluster. The proportion of patients who chose tonsillectomy was similar, irrespective of the patients' tonsillectomy preference score group (see Figure 22 Treatment chosen based on patient tonsillectomy preference score cluster grouping).

ENT surgeons' tonsillectomy preference scores are displayed in Figure 23 ENT surgeons' tonsillectomy preference scores. Calinski pseudo F score index (137.80) was greatest for 3 clusters, wheras Duda's Je(2)/Je(1) (0.30) index was greatest for 2 clusters. Silhouette score was marginally better for 3 clusters (3 clusters : 0.29 vs 2 cluster 0.15) and therefore 3 clusters were used – Figure 24 ENT surgeons' cluster groupings based on tonsillectomy preference scores. Table 31 shows the how the consultations are divided by ENT surgeons' cluster. It appeared that a greater proportion of patients who saw ENT surgeons' clustered with a stronger preference for tonsillectomy (83%) ended up choosing tonsillectomy compared to patients seeing ENT surgeons with a low tonsillectomy preference (60%) see Figure 25 Proportion of patients choosing tonsillectomy based on ENT surgeons' tonsillectomy preference cluster' tonsillectomy preference cluster. This was confirmed on multivariable logistic regression testing (Adjusted OR 3.88, 95%CI 1.01-14.97) – see Table 32 Multivariable model of predictors of treatment choice.

Chance of choosing tonsillectomy	Percentage who chose conservative (n)	Percentage who chose tonsillectomy (n)	Adjusted OR (95%CI)	P>z
Log patient uncertainty	36(20%)	148(80%)	0.42(0.19-0.95)	0.04
Log surgeon uncertainty	36(20%)	148(80%)	0.00(0.00-0.08)	0.01
Patient's age				
15-29	18% (24)	82% (109)	1	
30-59	24% (1)	76% (35)	0.71(0.23-2.22)	0.56

Patient's gender				
Male	16% (5)	84% (27)	1	
Female	20% (30)	79% (118)	0.94(0.24-3.72)	0.93
Patient's ethnicity				
White	17% (25)	82% (118)	1	
Non-white	27% (10)	73% (27)	0.51(0.15-1.71)	0.27
Self-reported episodes of tonsillitis in last 12 months				
0-5	22% (13)	78% (17)	1	
6-7	17% (10)	83% (49)	1.40(0.38-5.11)	0.32
>7	19% (12)	81% (50)	0.52(0.15-1.82)	
Self-reported days off from work due to tonsillitis in last 12 months				
0-5	25% (13)	75% (38)	1	
6-10	21% (10)	77% (37)	0.81(0.21-3.08)	0.6
>10	15% (12)	85% (67)	1.52(0.44-5.24)	
Type of hospital				
University	22% (13)	78% (47)	1	
District General	19% (23)	81% (101)	1.15(0.27-4.86)	0.85
Surgeon's age				
20-29	1(13%)	7(88%)	1	
30-39	11(19%)	47(81%)	0.24(0.01-4.94)	
40-49	16(26%)	45(74%)	0.40(0.01-11.02)	0.8
>50	8(16%)	43(84%)	0.41(0.01-13.04)	
Surgeon's sex				
Male	25(20%)	101(80%)	1	0.94

Female	11(21%)	41(79%)	0.94(0.23-3.89)	
Surgeon ethnicity				
White	16(19%)	70(81%)	1	
Non-white	20(22%)	71(78%)	1.29(0.37-4.49)	0.69
Surgeon's grade				
Consultant	21(23%)	72(77%)	1	
Registrar	8(18%)	36(82%)	2.48(0.31-19.74)	
Staff grade	4(27%)	11(73%)	0.67(0.10-4.35)	0.4
Associate Specialist	2(9%)	21(91%)	5.50(0.52-57.79)	
Core trainee	1(33%)	2(67%)	0.29(0.00-18.93)	

Table 28 Multilevel multivariable model of predictors of choosing tonsillectomy

This table shows associations between demographics and the probability of tonsillectomy (n=180).

Consistency ratio cut off	No of consultations	Tonsillectomies	Conservative therapy	Mean Tonsillectomy preference patient	Mean Tonsillectomy preference surgeon
0.4	170	139	31	0.56(0.55-0.57)	0.56(0.55-0.57)
0.3	140	115	25	0.56(0.54-0.57)	0.56(0.55-0.57)
0.2	90	78	12	0.56(0.54-0.58)	0.56(0.54-0.57)
0.1	20	18	2	0.59(0.55-0.63)	0.59(0.55-0.63)

Table 29 Sensitivity Analysis for consistency ratios

This table shows the mean tonsillectomy preferences for different consistency ratio cut offs, showing no relationship between outcome rankings and treatment chosen (n=160).

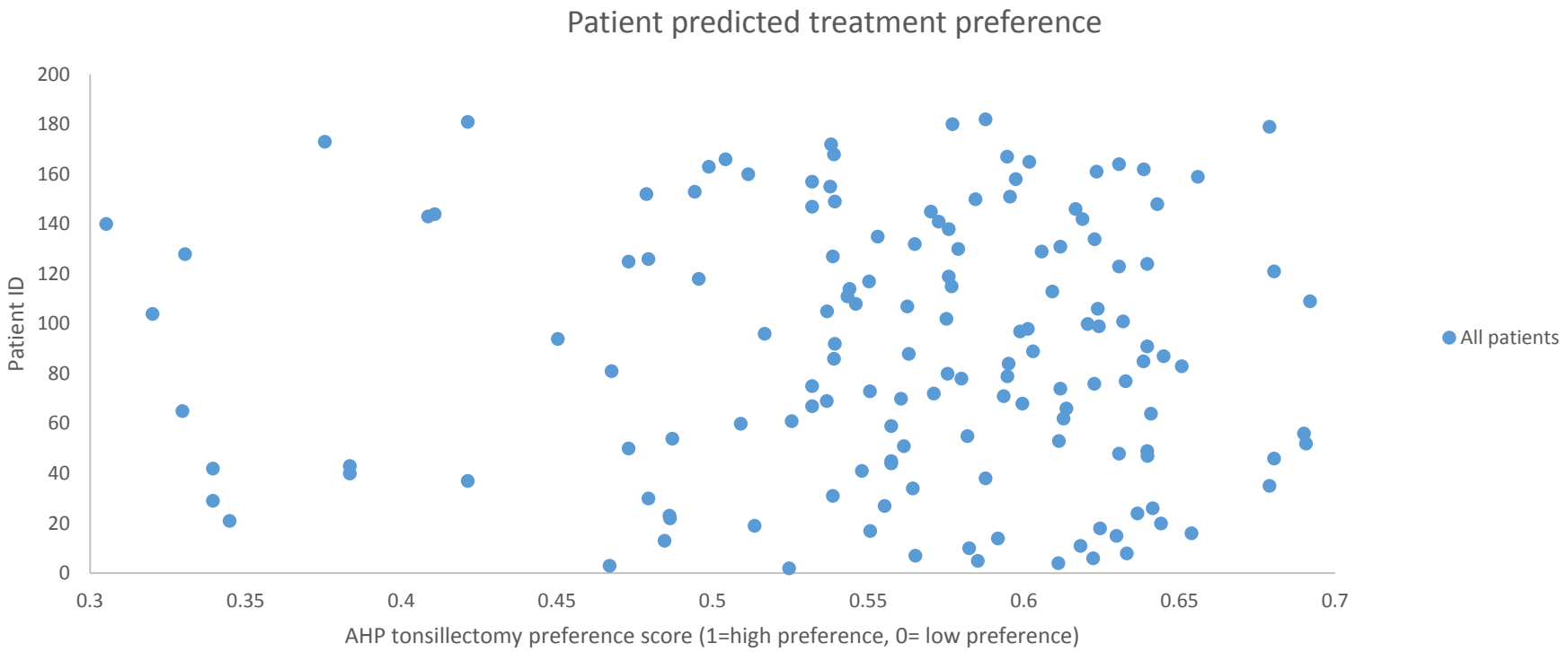


Figure 20 Patients' tonsillectomy preference scores

This is a graph that describes the tonsillectomy preference scores for patients, ranging from 0 (low preference) to 1 (high preference) for each patient in the study (n=167) This shows the distribution of preferences without cluster analysis. The y-axis is a periphrasis.

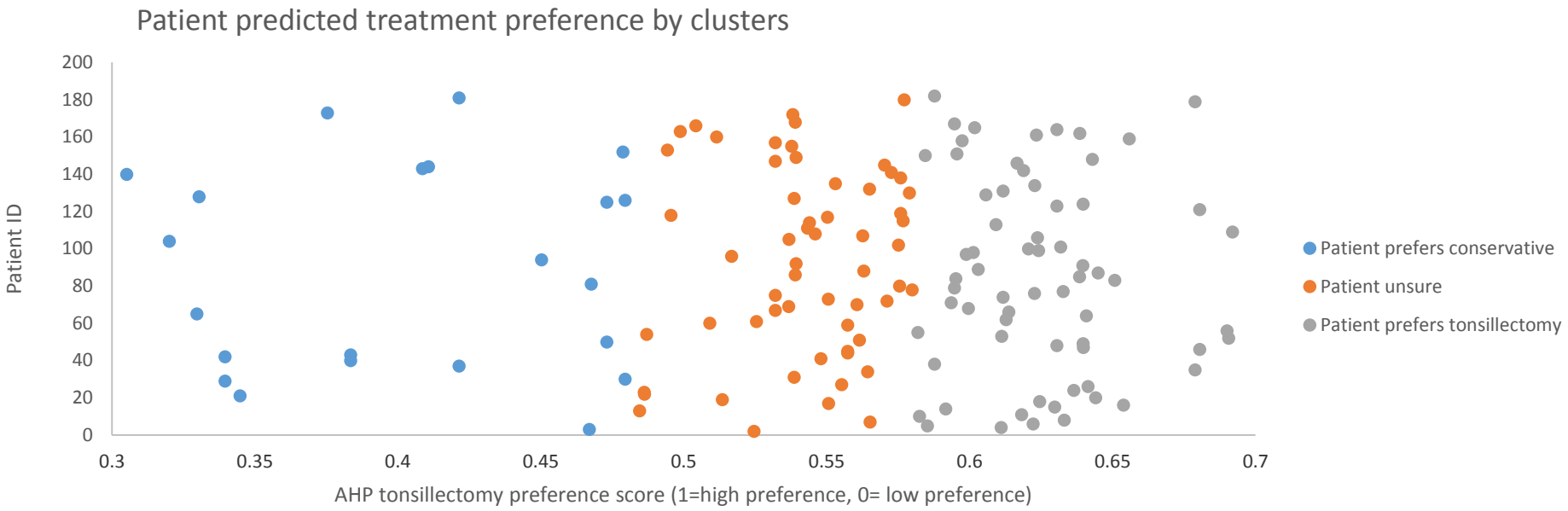


Figure 21 Cluster analysis of patient's tonsillectomy preference scores

This graph shows the results of the cluster analysis of patient tonsillectomy preference. Cluster analysis uses patient's data to create "preference profiles" which are used to create groups based on data, rather than requiring empirical categorisation using arbitrary thresholds. There are 22 patients belonging to the conservative cluster, 58 in the uncertain cluster and 66 in the tonsillectomy cluster. The y-axis is a periphrasis.

Cluster grouping	Mean Tonsillectomy preference score	Number of participants choosing tonsillectomy	Number of patients altogether	Odds ratio	P
Patient prefers conservative	0.40(0.38-0.43)	17 (77%)	22	1	0.85
Patient unsure	0.54(0.53-0.55)	46(80%)	58	1.18(0.31-4.41)	
Patient prefers tonsillectomy	0.63(0.62-0.63)	51(48%)	66	1.03(0.35-3.05)	

Table 30 Characteristics patient tonsillectomy preference clusters

This table shows the characteristics of the patient tonsillectomy clusters (n=146).

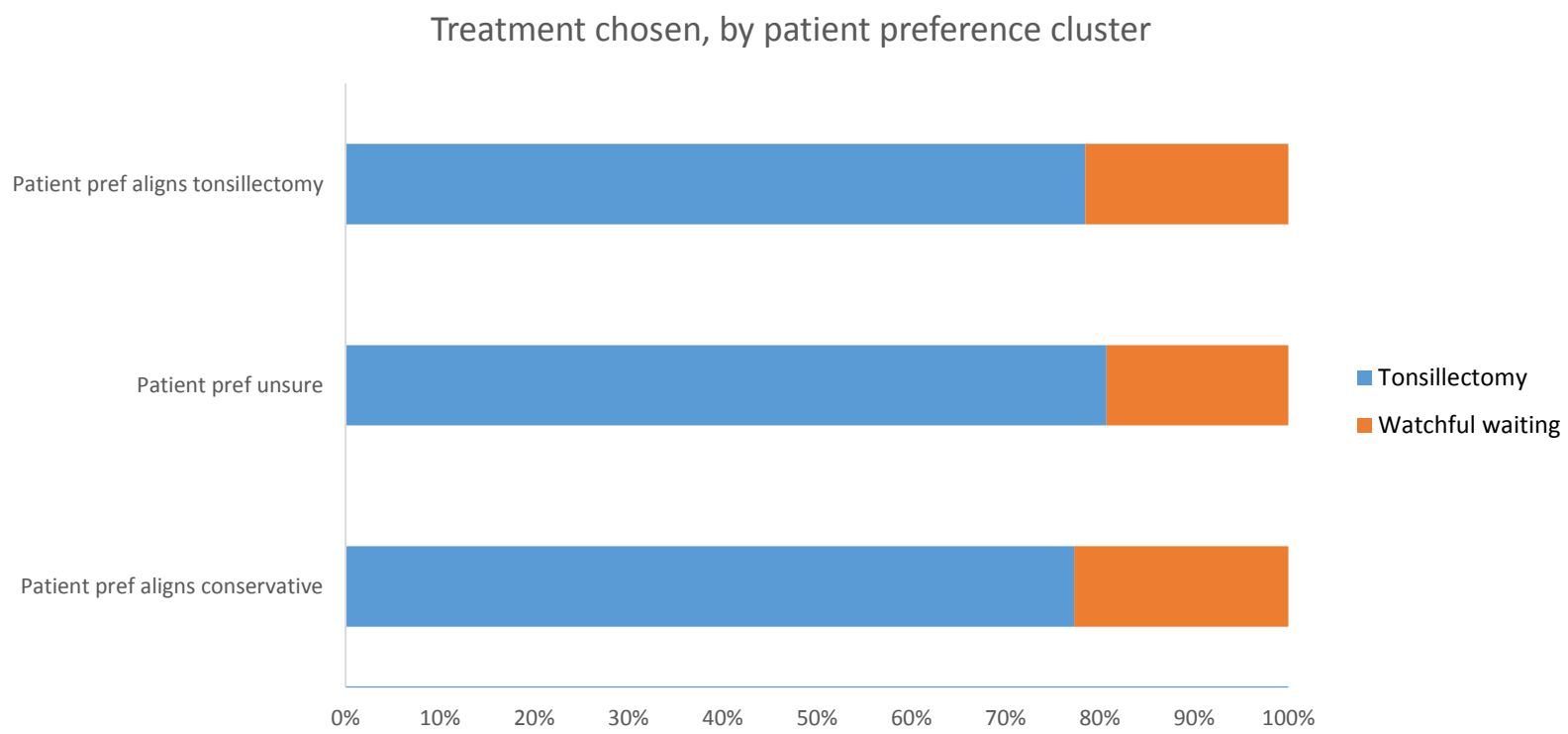


Figure 22 Treatment chosen based on patient tonsillectomy preference score cluster grouping

This figure illustrates the treatment that was chosen by each patient preference cluster, showing that in all groups the most likely result was tonsillectomy.

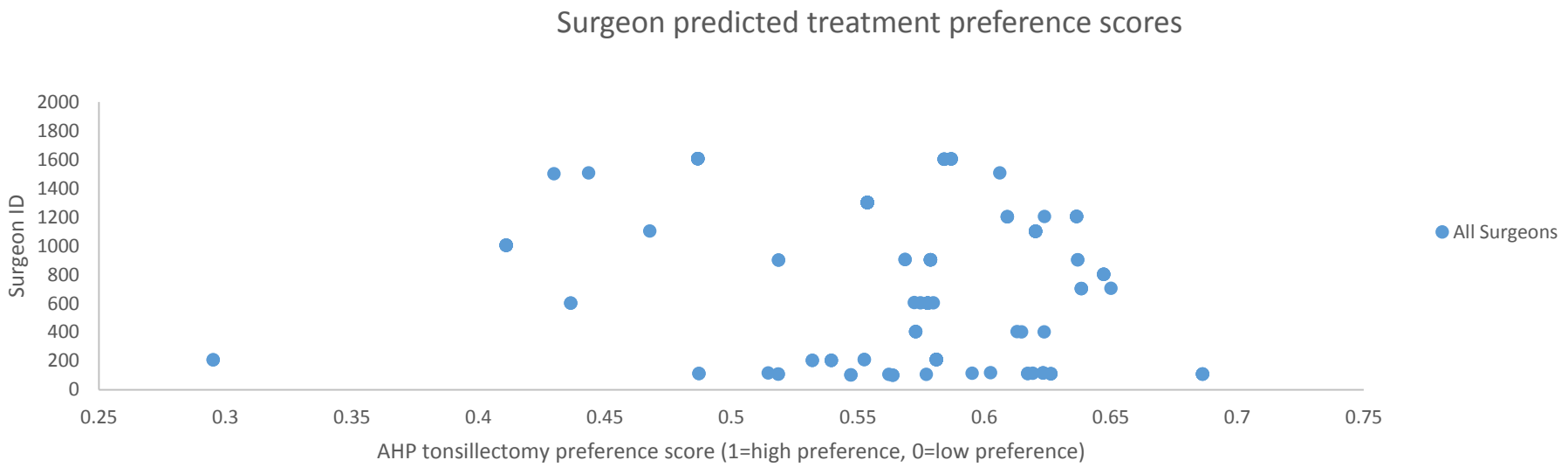


Figure 23 ENT surgeons' tonsillectomy preference scores

This is a graph that describes the tonsillectomy preference scores for surgeons, ranging from 0 (low preference) to 1 (high preference) for each patient in the study (n=167) This shows the distribution of preferences without cluster analysis. The y-axis is a periphrasis.

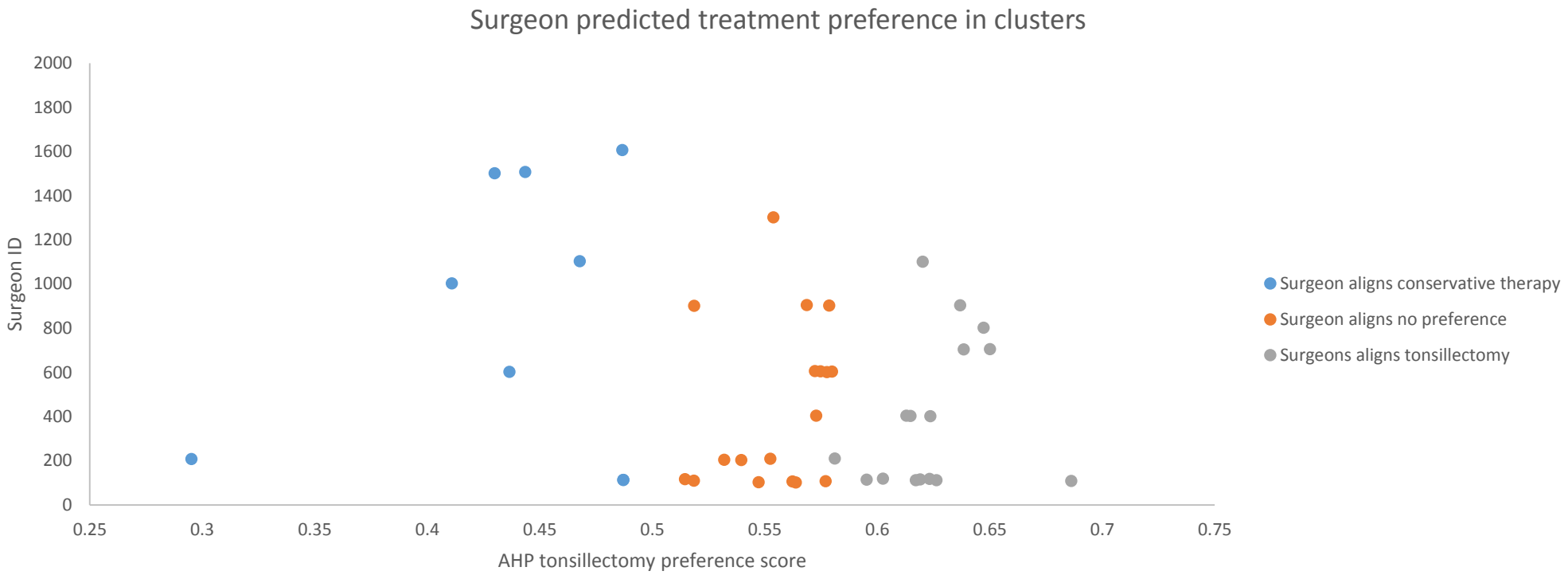


Figure 24 ENT surgeons' cluster groupings based on tonsillectomy preference scores

This graph shows the results of the cluster analysis of surgeon tonsillectomy preference. Cluster analysis uses surgeon's responses to create "preference profiles" which are used to create groups based on data, rather than requiring empirical categorisation using arbitrary thresholds. There are 12 surgeons belonging to the conservative cluster, 82 in the uncertain cluster and 73 in the tonsillectomy cluster. The y-axis is a periphrasis.

Cluster grouping	Mean Tonsillectomy preference score	Number of consultations ending in tonsillectomy	Number of consultation altogether	Odds ratio	P
Surgeon aligns conservative	0.38(0.35-0.41)	6(60%)	12	1	0.01
Surgeon aligns no preference	0.54(0.54-0.55)	62(81%)	82	2.95(1.25-6.95)	
Surgeon aligns tonsillectomy	0.62(0.61-0.63)	54 (83%)	73	3.27(1.33-8.04)	

Table 31 Characteristics of consultations based on ENT surgeon tonsillectomy preference cluster

This table shows the characteristics of the consultations, separated by their ENT surgeon tonsillectomy cluster, showing significant increase in likelihood of tonsillectomy if the surgeon's preference is for tonsillectomies (n=167).

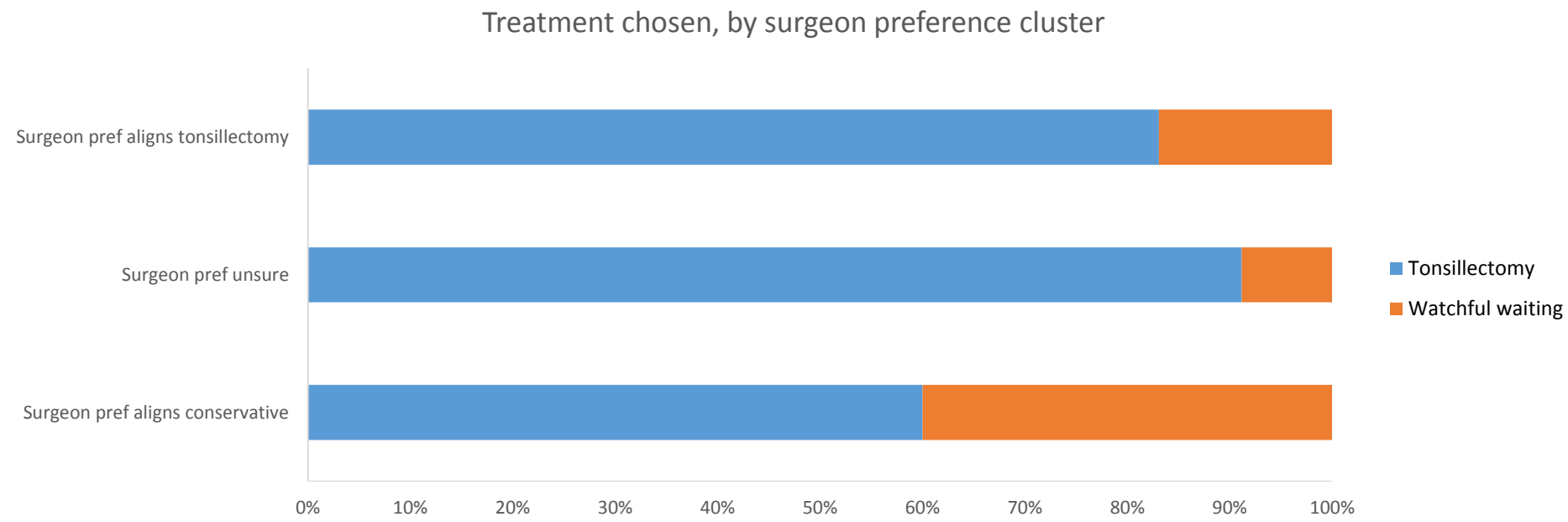


Figure 25 Proportion of patients choosing tonsillectomy based on ENT surgeons' tonsillectomy preference cluster' tonsillectomy preference cluster

.

Odds of choosing tonsillectomy	Adjusted Odds ratio (95% CI)	P-value
ENT surgeon tonsillectomy preference group		
Aligns conservative therapy	1	
Unsure	3.74(1.11-12.56)	0.01
Aligns Tonsillectomy	3.88(1.01-14.97)	
Patient tonsillectomy preference group		
Aligns conservative therapy	1	
Unsure	1.67(0.37-7.62)	0.65
Aligns Tonsillectomy	2.38(0.79-7.13)	
Patient age		
15-29	1	0.08
30-59	0.34(0.11-1.01)	
Patient gender		
Male	1	0.7
Female	0.79(0.29-2.13)	
Patient ethnicity		
White	1	0.45
Non-white	0.68(0.24-1.94)	
Self-reported episodes of tonsillitis in last 12 months		
0-5	1	
6-7	7.02(1.34-36.89)	0.06
>7	1.26(0.37-4.38)	
Type of hospital		
University	1	0.29
District general	1.82(0.57-5.84)	

Table 32 Multivariable model of predictors of treatment choice

This table shows the variables which may predict treatment choice alongside their odds ratio and p-value, illustrating the surgeons' preference is the major predictor of tonsillectomy.

Predictors of belonging to high ENT surgeons' tonsillectomy preference cluster	Adjusted RRR	P value
Hospital type		
University	1	0.02
District general	2.90(1.18-7.14)	
Surgeon age		
20-29	1	0.02
30-39	0.13(0.02-0.87)	
40-49	1.35(0.15-11.79)	
50-69	1.26(0.13-12.65)	
Surgeon gender		
Male	1	
Female	1.06(0.38-2.95)	0.92
Surgeon Ethnicity		
White	1	0.61
Non-white	0.80(0.34-1.87)	
Surgeon grade		
Consultant	1	0.71
Registrar	1.41(0.32-6.26)	
Staff Grade	2.27(0.40-12.79)	
Associate Specialist	0.39(0.12-1.33)	
Core Trainee	7.26(0.30-176.02)	

Table 33 Predictors of belonging to high tonsillectomy preference cluster for surgeons.

This table shows the predictors of belonging to a high tonsillectomy preference cluster, indicating that being in a DGH and age are the strongest predictors of tonsillectomy preference

Discussion

Summary

I have shown that patients' decisional uncertainty was generally low, suggesting they felt certain about the treatment they had decided on. ENT surgeons reported uncertainty scores that are considered clinically important (>25), suggesting they were uncertain, or conflicted, with regards to the treatment chosen. Both patients and ENT surgeons reported higher uncertainty scores when watchful waiting was chosen. This is understandable given that those who are unsure of which treatment to undergo may choose to delay their choice – i.e. undergo watchful waiting. I found no association between patient and ENT surgeons' uncertainty.

Interestingly, whilst patients' tonsillectomy preference scores were not related to treatment chosen, ENT surgeons' tonsillectomy preference scores were associated with the treatment chosen, with patients being nearly four times more likely to choose tonsillectomy if their surgeon had a high tonsillectomy preference score. This association remained even after controlling for patients' preferences and other patient and hospital characteristics considered to influence the treatment decision. This suggests that surgeons' pre-consultation characteristics have a greater role to play in the decision of whether tonsillectomy is chosen than any other patient or disease characteristics that may have been identified during the consultation. Hospital characteristics predicted surgeon tonsillectomy preference scores: Surgeons who worked in district general hospitals were almost three times more likely to report high PARTT scores than surgeons who worked in university hospitals. Whilst the reason for this is unclear, it may be due to the fact that teaching hospitals have a more complex workload when compared to DGHs. This may result in there being less theatre time dedicated to more routine procedures. Further research is needed to investigate this.

Results in the context of existing literature

Our patient DCS scores were of a similar magnitude to those described in other studies investigating preference sensitive surgical decisions such as male sterilisation (320), trigger finger treatment (321) and surgery for breast cancer(322). Our patients reported lower scores compared with patients considering mechanical ventilation in COPD (323) and

hormone replacement therapy when at risk of blood clots(324). Our ENT surgeons reported higher uncertainty scores (PDPAI) compared with surgeons managing patients with carpel tunnel syndrome(321), which may be reflective of the greater professional uncertainty that surrounds the management of recurring tonsillitis.

When choosing watchful waiting higher uncertainty scores were reported by both surgeons and patients. Whilst watchful waiting may have inherently higher uncertainty another possibility may relate to differences in the shared decision making (SDM) process. In fact, trials of patient decision aids designed to optimise shared decision making were reviewed by Cochrane (270) and found to show lower decisional conflict related to feeling unclear about personal values (difference -4.81 of 100; 95% -7.23 to -2.40) and lower decisional conflict related to feeling uninformed (difference -6.43 of 100; 95% -9.16 to -3.70). It is possible that the higher uncertainty scores relate to reduced or ineffective shared decision making during the consultation. Alternatively, patients who choose conservative therapy may be more appreciative of the risks and benefits of their choice and thus report a higher uncertainty score (325). Another explanation could be related to cognitive dissonance. That is, a potential mechanism for lower uncertainty scores in patients who choose tonsillectomy, is that they may have adjusted their uncertainty score post hoc to justify the decision they just made (327). Qualitative studies would help investigate the implications of our findings.

Our results suggest that patients are more likely to choose treatments based on surgeon characteristics, which were established prior to their consultation, rather than their own treatment preferences or their disease severity. In part this may be related deferring decisions to experts. Qualitative studies of shared decision making that have shown patients see their surgeons as occupying expert roles and frequently defer decisions to them(328). In shared-decision making, it is the role of the surgeon to elicit patient preferences and help them choose which treatment will provide outcomes that the patient values. Surgeon PARTT scores elicited refer to surgeon treatment preference in the context of the medical consultation. The surgeon PARTT score may be driven by surgeon's - evaluation of tonsillectomy effectiveness; perception of what outcomes the patient values; proxy bias; and personal tonsillectomy preference. There is already a literature regarding how the surgeon's implicit perception of

patient's preferences differ substantially from the patients' actual preferences(253,329,330).

Our results may help explain previous findings from studies more specific to the management of recurring tonsillitis, albeit in children. These studies report that ENT surgeons' treatment preference are more deterministic of treatment chosen than the patients' disease characteristics (22,23) and that variations in ENT surgeons' consultation styles systematically predispose to a treatment being selected (17). Both observations may be a result of latent surgeons' preferences towards a treatment. However, these studies were conducted in an era when there was little evidence on the effectiveness of tonsillectomy, no national guidelines and paternalistic surgeon decision making, all of which would have magnified regional tonsillectomy rate variation.

Whilst it is surprising that ENT surgeons' treatment preference, in the context of the medical consultation, can still be so influential, despite increased patient autonomy, advances in psychology over the last four decades have shown that even subtle differences in medical consultations can affect treatment choice. Indeed the investigations of framing bias have shown that patients are influenced into selecting a treatment depending on how outcome information is 'framed' by their surgeon: Patients select the risk averse option more frequently when the same information is conveyed in terms of harm rather than benefit(331) (193). Psychological theory suggests that the format in which an argument is framed occurs subconsciously and is a consequence of the beliefs and values held by people presenting the information (332). Alternatively, our findings may suggest how the shared decision making process can be more complex than just sharing evidence on treatment benefits and risks and eliciting patients' treatment preferences. Patients may not feel equipped, and may not want to make a decision and may defer the decision to the expertise of the surgeon. A systematic review of studies that measured patients' decision role preferences showed that 21% of all patients preferred to delegate decision making roles to their care givers (333). This number dropped to 14% when the authors analysed decisions about surgery only. Additionally, the authors reported that the proportion of patients who wish to delegate their medical decisions from 43% in the 1970's to less than 16% between 2000-2007. Whilst the delegation of a decision is a reasonable response it exposes the

patient to variation in surgeon treatment preferences, with two similar patients being offered different treatments based on the surgeon they saw more than their personal values or their disease severity. And this response may be associated with the type of hospital the patient visited. Qualitative work focusing on the consultation may help elucidate how ENT surgeons' a priori preferences affect the risk of receiving a treatment and whether it was undertaken purposefully, due to patient decision paralysis, or unconsciously through communication biases.

Strengths

Overall, this the largest observational study of decision making in adults with recurring tonsillitis using a representative sample. The study measured constructs that had historically been associated with regional surgical rate variation, that is decisional uncertainty (as a potential surrogate marker of professional uncertainty) and treatment preference in both ENT surgeons and patients. PARTT is the only measure available that elicits preferences in adults with recurring tonsillitis and was designed using robust methodology that incorporated available evidence (using systematic review and critical appraisal) with key-stakeholder views (thematic review of focus groups of adults with recurring tonsillitis). AHP is the only preference measurement method that can reliably quantify inconsistency. Other methods qualitatively judge inconsistency giving the investigators the option to include or exclude inconsistent responses- with little transparency to the manner this decision is made.

PARTT gave little inconsistency in our population compared to other healthcare preference studies using AHP measurement methods(334). I used sensitivity analyses to transparently assess the impact of inconsistent comparisons (consistency ratio 0.2-0.5) and found they made no difference to participants' rankings, therefore could include them in our analyses. This suggests that values may not always be well formed and inconsistency should not immediately be used to define lack of validity, as has been found in other AHP studies (335,336).

The sample was representative; ENT surgeons recruited were similar to membership profiles of the national ENT Association (ENTUK); Patients were similar in age, sex and ethnicity to

patients undergoing tonsillectomy in our CPRD-HES database and in the National Prospective Tonsillectomy Audit. Given the simplicity of the study, recruitment rate was very high and missing data very low. Patients and ENT surgeons completed their questionnaires immediately following the consultation, preventing recall bias affecting our scores. I could recruit ENT surgeons from all grades and therefore, control for this potential confounding variable in our final model.

Limitations

More of our sites were teaching hospitals (29%) compared to the national proportion (3%), however, there is no evidence to suggest that recurring tonsillitis patients (in patients for emergency admissions or in patients for elective referrals) are different between district general hospitals and teaching hospitals. The cross-sectional nature of the study prevents deeper understanding of the causal interactions between decisional uncertainty and treatment chosen.

Further limitations of our study can be discussed under two main topics: First, the limitations of the instruments used. Second, the power of the sample size to reach the conclusions.

Potential limitations of instrument used

Whilst the DCS has been well validated across several clinical conditions, this is its first use in recurring tonsillitis. Therefore, I undertook a limited evaluation of its psychometric properties in our sample. The DCS had high internal consistency (Cronbach's alpha= 0.96), good item-total correlations (0.68-0.83), however, item analysis suggested a floor effect for most items. Improved scale response categories may have allowed a more detailed understanding of the relationship between treatment and uncertainty score. However, it was not possible to evaluate validity or responsiveness in our study.

PARTT showed that not all respondents were consistent in the manner they completed their pairwise comparisons. 13% of patients (23/174) and 16% of ENT surgeons (10/62) were inconsistent in the manner they recorded their priorities (consistency ratio>0.2). This may reflect an inherent weakness in our measurement process. There are five main reasons described for inconsistent responses in AHP questionnaires: Clerical error; use of extreme values; poor model structure; lack of understanding and true intransivity (337). My results

show that the use of extreme values was more common in those who showed inconsistent response (consistency ratio >0.2) compared to those who did not. It may have been that these participants were using extreme values to support the direction of their preference as opposed to strength of their preference e.g. a respondent chooses *extreme preference* of reducing visits to the GP over reducing days of sore throat and then chooses *extreme preference* of reducing days of sore throat over reducing halitosis. Whilst AHP can be used to understand that this respondent's ranked outcome priorities are reducing visits to the GP (first), followed by reducing days of sore throat (second), followed by reducing halitosis (third), it is difficult to ascertain the strength of the preference between the three outcomes (is reducing visits to the GP two or three times as important reducing days of sore throat?).

Previous AHP studies have dealt with inconsistent reporting in different ways. One way is to report inconsistency back to the respondent in real time and allow them to change their response. However, in this study I did not want to place further burden on patients. Others have addressed inconsistency by excluding participants from the analysis who were inconsistent(338-341). However, I felt that this would unnecessarily exclude patients with valid responses. I excluded all those who had very high inconsistency from analysis (consistency ratio>0.5) and ran sensitivity analyses on those with moderate degrees of inconsistency to see if it changed the treatment preference scores for the group. Since there were no differences in the overall tonsillectomy preference scores, or the ranked outcome priorities I did not exclude those with moderate inconsistency (consistency ratio 0.2-0.5).

In addition to these observed reasons for inconsistency, a range of publications discuss the AHP scale and suggest that the scale itself has limitations (e.g., being bounded, not continuous, not representing verbal judgments well, or not delivering balanced judgments) (342-346). Alternative scales avoiding this and other potential weaknesses of the AHP scale have been proposed (e.g., Lootsma or other geometric scales; Ji's derived transitive scale; different continuous, smaller, or wider scales) and are still being discussed (343,345-348)]. Future studies might investigate the extent to which the chosen scale contributes to observed inconsistencies and the role inconsistency feedback to the respondent has on the results.

Sample size

Although I recruited more patients than expected, fewer patients chose conservative therapy than anticipated (20% versus 33%). Power analysis of our DCS results showed I had sufficient participants to detect an effect as small as 0.42, and since the effects I demonstrated were larger I feel justified in my conclusions. I also conducted a post-hoc power analysis of our main preference finding (i.e. ENT surgeon PARTT score is associated with treatment chosen). I used a two tailed Chi squared test with a 5% significance level based on the observed effect (adjusted OR 6.34), the sample size between the two surgeon preference groups (n=90) and an allocation ratio of 1, to show that I was sufficiently powered (power=81%)(349). Therefore, the study was adequately powered to report this finding.

I repeated this test to examine our finding that there was no relationship between patient tonsillectomy preference and treatment chosen. My results showed that the study was adequately powered to reach this conclusion (Chi squared showed that 21000 patients would need to be recruited to detect (with 80% chance) this observed effect (adjusted odds ratio 0.94) as significant at the 5% level).

Conclusions

Our results suggest patients' preferences appear to play little role in the treatment decision. Conversely, ENT surgeons' implicit treatment preferences seem to have a greater impact on treatment decisions, especially when the surgeon aligns to conservative therapy. This observation seems to remain true even after accounting for patients' preferences and disease severity variables. Whilst it can be expected that surgeons develop an implicit preference for treatment of adults with recurring tonsillitis based on their personal experience and the experience of others (local working environment as well as national and internationally published literature), it becomes an issue when there is considerable variation between surgeons.

CHAPTER EIGHT: Discussion

Chapter synopsis

In this final chapter I have discussed the key findings of my thesis in the context of the current literature and overall strengths and weaknesses of this body of work. I have discussed potential avenues for future research, including qualitative studies to better understand my findings in relation to surgeons' proxy rating of patient preference. I have concluded with implications for national policy, which include targeting interventions in the community as well as using patient decision aids to improve shared decision making.

Key findings in the context of current evidence

There are three key findings of my thesis. Firstly, regional tonsillectomy rate variations reflect regional variations in 'need' of the population; Second, regional tonsillectomy rate variations are greater for children than adults. And finally, treatment decisions for adults with recurring tonsillitis are more influenced by surgeon's treatment preferences than patient preferences or severity.

Regional tonsillectomy rate variations reflect regional variations in 'need' of populations

My studies have shown that there is regional variation in the 'need' for tonsillectomy by demonstrating regional variations in recurring sore throat in the primary care and self-reported sore throat in the community. After accounting for regional population characteristics related to 'need' for tonsillectomy, I found that the regional disparity was considerably lower than originally published for children (SCV=2 vs 8.4 (20)), suggesting that a large part of the original variation described relates to the population 'need'.

As described in Chapter 1, the term 'need' has been defined as the population who could benefit from a treatment (60). Variation in surgical 'need' has previously been described in relation to:

1. Regional variations in disease incidence (which may reflect local demographics and lifestyle factors)
2. Regional variations in disease detection (which may reflect local access to services and help seeking behaviour such as GP consultation)
3. Regional variations in patient treatment preference (which may reflect local social norms).

In relation to tonsillectomy 'need' could either be approximated by:

1. Sore throat incidence in the community.
2. Recurring sore throat detection in primary care
3. Patient treatment preference measurements in secondary care.

Whilst there are a handful of studies that report the incidence of sore throat in the community (350), consultation rate for sore throat(2,142,143) and recurring sore throat (139) in

primary care, recurring tonsillitis (351) across settings, there are no studies to date that have investigated regional variations in sore throat from the community to primary care, let alone patient treatment preferences for recurring tonsillitis. However, my findings support findings in other clinical areas which have demonstrated that the community incidence of acute respiratory infections (352) (353) (354,355), GP consultation for acute upper respiratory tract infections (69) (70-72) (73,75,76) and patients willingness to undergo surgery (77,78,83,90,91) are all affected by population characteristics that are known to vary across regions. Population characteristics also seem to predict help seeking behaviour more generally for urgent care in primary (356) and secondary healthcare centres (357).

For the first time, I have been able to approximate variations in 'need' by reporting regional disparities in incidence of community sore throat and detection rates of recurring sore throat in secondary care. My findings are strengthened by re-analysis of my results which demonstrate similar patterns of variation across all health care settings, initially captured on completely independent databases. For example, residents of East Midlands had amongst the highest rates of tonsillectomy, as undertaken in secondary care and captured on the HES database; this observation appears in part to be explained by the same residents having amongst the highest rates of primary care consultations for recurring sore throat, as captured on the CPRD database, and self-reported sore throats in the community, as captured on the FluWatch database. Bringing this together, my studies suggest that a large component of regional tonsillectomy rate variation is generated prior to secondary care visits.

Indeed, much of the regional variation in tonsillectomy rates may occur even prior to primary care attendance for recurring sore throat. My finding that much of regional variation in the primary care detection of recurring tonsillitis becomes non-significant once regional population characteristics (e.g. age, sex, ethnicity, socioeconomic status etc.) are taken into account (SCV=2 for both adults and children). This further demonstrates that regional population characteristics, independent of regional health care variables, contribute more significantly to regional variations in recurring tonsillitis detection and management than previously described. This may, in part, help explain why interventions aimed at healthcare professionals have resulted in minimal change in the observed regional tonsillectomy rate variation.

Regional variations are greater for children than for adults.

Whilst regional tonsillectomy rate variations in children are well documented, there are no studies, to date, that report on those rate variations in adults despite 40% of all tonsillectomies being performed in adults (HES database 1997-2001). My study showed that regional tonsillectomy rates varied up to 20% for adults but up to 300% for children. To better understand the drivers of this disparity I investigated overall patterns of sore throat from the community through to secondary care separately for children and adults, at available time points (see Table 34 Incidence and rates of sore throat from community, through recurrence in primary care to tonsillectomy in secondary care).

	Self-reported sore throat (episodes/1000pt-years) Flu-Watch	Sore throat consultation rate in primary care (consultations/ 1000/pt-years)	Recurring sore throat (3 episodes in less than 12 months) (patients/ 1000-pt-years)	Tonsillectomy rate (patients/1000 pt-years)
0-4 years old	1168	250	104	0.6
5-15 years old	1679	142	66	2.5
16-24 years old	1559	116	2.9	1.8
25-44 years old	2000	80	2.6	0.5

Table 34 Incidence and rates of sore throat from community, through recurrence in primary care to tonsillectomy in secondary care

This table shows the incidence and rates of sore throat in the community in different age groups, showing the progression through the healthcare system.

Table 34 demonstrates that whilst the incidence of self-reported sore throat was similar between ages 5-15 and 16-24, the proportion of those who consulted for sore throat (first versus second column) was much greater in children (250/1168 patient years) compared to

young adults (116/1159 patient years). A greater proportion of children who see their GP once with sore throat go on to have recurring sore throat (66/142 patient years) compared to young adults (2.9/116 patient years). This observation may be explained by qualitative studies that have shown parents' decision to bring their children to the GP with acute respiratory illness is influenced by a greater perception of threat severity and increased expectation of assessment, information, advice or treatment(69-72,168) compared to similar severity levels of acute sore throat (167) or respiratory tract infections (169) in adults. Therefore, it seems that children are more likely to attend their GP for varying levels of disease severity (as observed in my data), whereas adults are more likely to attend for higher levels of disease severity. There may be consensus amongst GPs that adults who attend repeatedly for sore throat infections are more likely to have severe disease and benefit from referral for tonsillectomy. Professional consensus in the management of adults would reduce regional variation in this group. This has never been directly investigated, but can be inferred from my results: Many adults (16-24years old) who attended primary care for recurring sore throats ended up having a tonsillectomy (62%), whereas only 4% of children (5-15 years old) had a tonsillectomy following the same number of consultations. These results may reflect strongly held consensus amongst the medical community of the benefit of tonsillectomy in the adult population, which has the effect of reducing regional variation.

Surgeon's treatment preferences influence treatment decisions for adults with recurring tonsillitis

In my study of decision making, I asked surgeons to complete PARTT once, from the perspective of an adult patient with recurring tonsillitis, before they saw the first study patient at their site. I found that surgeons who had implicit preference scores that tended towards conservative treatment of recurring tonsillitis were 74% less likely to list their patients for a tonsillectomy than those whose had an implicit preference towards tonsillectomy; this preference was independent of the patients' preferences (measured using PARTT) or markers of disease severity (e.g. episodes in last year, time off routine activities etc.). Surgeons whose PARTT scores tended towards a non-surgical strategy placed a higher value on reducing the

risk of bleeding compared to surgeons favouring surgery. My study did not have the power to investigate if surgeons' treatment preference scores varied geographically; it did however show that surgeons' treatment preference scores varied with the type of hospital s/he worked in. Perhaps due to less capacity for tonsillectomies in teaching hospitals when compared with DGHs, as there is more alternative work in tertiary care.

There is no study to date that examines treatment decision making in adults with recurring tonsillitis and therefore I cannot make direct comparisons. In children with recurring tonsillitis (n=400), Bloor's (23) studied treatment decision making and found that there was considerable variation in the management of this condition by ENT surgeons.

Bloor did not directly elicit surgeon treatment preference, but observed that some surgeons tended towards tonsillectomy and others watchful waiting, irrespective of patient or disease variables{Bloor:1976bm, Bloor:1978vy}. Whilst it is to be expected that surgeons develop a personal treatment preference over time based on their personal or local experiences, it becomes problematic when patients with the same condition severity and personal treatment preference are exposed to different treatments by different surgeons.

Strengths and Limitations

The work presented in my thesis is a unique mixed methods study that explores surgical rate variation. The work presented is the first time that health care data analysis has been used across three health care settings from the community to secondary care to better understand regional surgical rate variations and report on findings that have traditionally been missed by previous studies since they only investigated variations across one setting. I developed a new AHP instrument capable of efficiently eliciting treatment preferences for adults with recurring tonsillitis, based on a well validated process (AHP) and created using systematic reviews and critical appraisal of available evidence regarding treatment outcomes, and thematic analysis of focus groups to define outcomes important to patients. I used multiple methods to develop a better understanding of decision making, including conceptual framework mapping, critical appraisal of patient reported outcome measures, design and deployment of a multi-centre national observational study of patient and surgeon decisional uncertainty and preference for the treatment of recurring tonsillitis. My findings into decision making reveal

the complexity of the process and difficulty in measuring it, the difficulty inherent to preference elicitation and the importance of assessing this process from both the patient's and surgeon's perspective, which would have been difficult to ascertain in the absence of robust and varied methods.

Previous studies have alluded to the 'surgical signature' where consultations with surgeons seem to result in one treatment more frequently being chosen compared to another. This has frequently been described as surgeon preference. However, surgeon preference in the clinical encounter, is a complicated process that involves beliefs around perceived treatment efficacy being combined with beliefs around which outcomes are important to their patients, which may be further complicated with their personal preferences. Whilst I could have assessed surgeon treatment preference by asking surgeons to complete a visual analogue scale from 0-10 based on their tonsillectomy preference, it would have missed the complexity of this construct. By asking the surgeon to complete a robust preference elicitation instrument from the perspective of a typical recurring tonsillitis patient I could capture the complexity of this construct without diminishing it.

However, whilst this may be a strength, there is also a limitation of this method: There is the difficulty in interpreting the results. Surgeon PARTT scores, completed from a hypothetical patient's perspective, seem to have more influence on the treatment decision than the real patient's treatment scores. It is difficult to unpick whether this relates to 1. The surgeons' concept of patient important outcomes 2. How effective the surgeon believed tonsillectomy was with respect to those outcomes 3. The surgeons' personal preferences or 4. Proxy bias introduced by asking surgeons to complete scores on behalf of someone else.

A further limitation of my thesis includes the use of only adults in the decision analysis study. As shown by the epidemiological investigation, regional tonsillectomy rate variation was greater for children than adults, and so analysing decision making for children would have been very useful. However, decisions between two agents, when one agent is acting as proxy for a third agent would have been difficult to interpret and so I chose, at least in the first instance, to investigate decisions solely between two agents.

Since tonsillectomy is a surgical procedure that can be used for both recurring tonsillitis and obstructive sleep apnoea, it could be considered a limitation to have not also investigated the variations in tonsillectomy rates for patients with obstructive sleep apnoea (OSA). This would have provided a more complete picture of all tonsillectomies. However, tracking this population through healthcare settings is extremely difficult given the poor coding for this population in primary care, in part related to diagnosis of this condition by multiple health specialists such as paediatricians and respiratory physicians. Given that OSA is a far less frequent reason for tonsillectomy compared to recurring tonsillitis, and the difficulties of defining the OSA denominator from the available data I felt it was justified to focus on recurring tonsillitis only. However, with OSA becoming a more frequent indication for tonsillectomy, care should be taken in devising policy based solely on recurring tonsillitis related tonsillectomy variation.

Whilst having several limitations, my thesis has demonstrated a robust mixed methods framework for investigating regional surgical rate variations. Which in relation to tonsillectomy I have shown are complicated by regional variations in the 'need' and potentially even influenced by surgeons during the decision-making process. This demonstrates the complexity of the issue that previously has been reduced to a simple description of 'surgical signature'.

Implications for future research

Developing the evidence base on patient relevant outcomes could help reduce professional uncertainty in the management of recurring tonsillitis in adults. Systematic review undertaken to design PARTT revealed a paucity of evidence on key outcomes related to adult tonsillectomy. Current research is already being undertaken to further detail the efficacy of this procedure (55).

Whilst I have been able to investigate tonsillectomy rate variations between 10 health districts, the data available to me, due to costs and information governance restrictions, did not allow me to look at variations between smaller health regions such as primary care trusts. This would have been valuable since most health policy is instigated at primary care trust level

and it would have allowed a more direct examination of the impact of local health policy on tonsillectomy rates. So future work should investigate variation at these levels to better help understand how local policy influences tonsillectomy rates.

Whilst I could show that my surgeon proxy ratings of patient treatment preference considerably influenced patients' treatments decisions, the complexity of the construct measure makes interpretation difficult as discussed above. Understanding whether surgeon treatment preference is based on what surgeons perceive is important to their patients, personal preferences, proxy bias or their perception of efficacy of tonsillectomy would help devise strategies to align treatment decisions to patients' treatment preferences. So, future work should investigate the drivers of surgeon treatment preference, using qualitative analysis of consultations and surgeon interviews.

It has been shown that effective integration of patient preferences into shared decision making has been associated with positive findings in other patient groups(358,359). My study showed that patient preferences may not be accurately integrated into the shared decision making as much as desired; the elicitation of those preferences may be a first step towards integrating them. In this respect PARTT may be able to play a part in elicitation of personal patient treatment preferences. Future work will investigate how PARTT can inform a patient decision aid, designed to help patients with recurring tonsillitis. Additionally, its role in feedback to surgeons about their latent preferences may help reduce their impact on the decision outcome. Future work investigating this may provide insights into the translation potential for a preference elicitation in shared decision making, both for the patient and the surgeon.

My thesis used tonsillectomy as an exemplar to investigate regional surgical rate variations. These methodologies could be applied to other conditions: 1) using community surveys to understand true variations in disease burden and their relationship to variations in treatment levels. 2) Using detailed electronic health records to more effectively adjust for the wide range of population factors that may affect variation – this is important to prevent wasted effort in attempts to reduce variation. 3) analysis of variation stratified by different age groups (attempts to reduce variation may ignore the fact that in some age groups where there is

minimal variation), 4) more detailed examination of the role of clinician vs patient preference in decision making.

Implications for policy

In a more general sense, there is a strong culture within the NHS of addressing variations of all kinds as a means of increasing quality and decreasing cost. In fact, the current government policy is to reduce unwarranted healthcare practice variations to reduce the variations in health outcomes(178). There are currently metrics of variation across almost every aspect of care (e.g. of Quality Outcome Frameworks, antibiotic prescribing, delays in cancer diagnosis, cancer outcomes, cancer screening uptake), however few of these account for patient characteristics to the extent that this thesis has, meaning that the initiatives may be a waste of effort at best and harmful at worst.

My study sheds light on what variation is warranted and provides a plausible reason as to why the policies to reduce tonsillectomy rate variations may have failed: High rates of tonsillectomy in certain regions are related to high rates of recurring sore throats in that region. Neither policy nor guidelines were directed at reducing the underlying disease burden but rather at aligning medical decision making and so may have failed to address the main driver of variation. Our study also shows that the there are many life style factors like smoking, population density (possibly acting through access to care) and obesity that increase the risk of recurring sore throats. Policy directed at reducing these may have a much greater effect on regional variation. Whilst the rate of recurring tonsillitis remains high in certain regions, tonsillectomy may be a cost-effective procedure (incremental cost per quality-adjusted life-year for paediatric tonsillectomy in England ranged from £3129 to £6904 per QALY gained (3) and the current problem may be underprovision rather than overprovision.

My observational study into decision making demonstrated that some patients are being exposed to different treatments depending on which surgeon they visit. Regardless of whether this results in regional tonsillectomy rate variation or not, it suggests that patients

with the same condition may end up with different treatments based on who they see during their consultation.

There are 3 potential strategies that could improve shared decision making: 1. Improve the evidence on the outcomes of treatment; 2. Implement strategies that reduce the impact of surgeon's proxy rating of patient preference on decision 3. Increase the effect of patient preference.

Firstly, it is not only important to increase the quantity and quality of evidence but also to re-prioritise studies so that we have information available regarding outcomes that matter to patients. In developing PARTT I have noted large disparities between outcomes that patients considered important and available studies about those outcomes. Once that information is available it will be key to make that knowledge more easily accessible to health care professionals and can be integrated into treatment decisions. When this is achieved successfully it has been shown to reduce bias in the stages of gathering evidence and choosing treatments(360-362). There are already initiatives like Core Outcome Measures in Effectiveness Trials (COMET) in existence that have started collating core outcome sets for treatments.

Secondly, surgeons' decision making could be improved by educating them about the existence of their biases, on the assumption that an awareness of the biases will permit them to avoid being influenced by them (363). Although informing people about biases is not very effective at reducing biased reasoning in non-medical settings (364), there is some indication that it can improve surgeons' reasoning in some respects. For example, Gruppen (326)found that informing surgeons about a specific bias — the tendency to be influenced by personal experience of how effective a treatment was in evaluating future treatment decisions — reduced its effect. This strategy is not only logical, but it is also justified by research showing that continuing medical education in general tends to improve surgeon performance across a variety of domains and outcomes (365). Policy could be directed to re-educating surgeons and informing them of their implicit bias towards a treatment, however, this would need to be done at a national level if it was to have any impact on regional tonsillectomy rate variation.

Finally, the impact of patient preference on treatment choice could be improved by formalising the process of patient preference elicitation and make it an explicit part of the shared decision making process. Policy could be directed at empowering patients in their treatment decisions. Whilst a recent study has shown more patients want an active part in their treatment decision than they did previously(333), there is no currently established framework to allow this. The value of understanding and using patient preferences in health care is well recognised (366,367). However, eliciting patient preferences is a difficult task that involves imagining a future health state, the likelihood of that health state and then the desirability of that health state. Finally, patients need to be able to compare amongst potential health states. It can be difficult for patients to understand uncertainty and risk or abstract constructs such as values and preferences. Additionally, attempting to compare values for potential future health states in the clinical consultation may stress the patient's decision making to an even greater extent. PARTT could elicit preferences using simple comparisons, but it remains unclear how well treatment preference had been formed prior to the use of the instrument.

Whilst there is evidence that shows skilled interpersonal interactions can elicit patient preferences efficiently(368,369), the time limited nature of the contemporary consultation rarely allows the opportunity to conduct the intense, interpersonal exploration required for preference elicitation. Patients in England are exposed to the second shortest consultation times across 10 developed countries(370). When preference elicitation is undertaken as an iterative, and deliberate process it can help a person understand and clarify personal values, health care situations, treatment options, and likely outcomes(371).Benefiting from behavioural decision-making research, an interactive analysis process is used to help an individual focus on key components. Whilst historically preference elicitation has been conducted by a skilled interviewer using probes and reflection, newer interactive computer systems have shown some success when used in addition to a human analyst (372).

Technology can assist in meeting the challenges inherent in eliciting and incorporating patient preferences with high value evidence in routine health care practice. Software developed to focus on elicitation and values clarification may help patients think hard about complex, abstract issues, such as the desirability of future states. A number of different approaches

have been trialled and are collectively termed patient decision aids(373-376). Systematic review of randomised controlled trials that examine the effects of patient decision aids has shown that they increase patient knowledge, improve patient perception of risk (relative risk 1.82, 95% CI 1.52-2.16), and improve congruency between stated values and treatment chosen (RR 1.51, 95% CI 1.17-1.97) (270). Additionally, the systematic review that I undertook showed that they also have the potential to change regional surgical rates(100).

A previous study that investigated the influence of the match between patients' preferences for information vs. information received. Authors reported that the trainee and consultant surgeons rated patients as adjusting better during surgery when the information provided before surgery matched patients preferences for information (377). Patients were also found to experience less anxiety and be more adaptive during an invasive medical procedure (catheterization) when provided with information matched to patients' preferences (378). Patients' adjusted better during surgery, had lower self-reported pain and reported better satisfaction when their preferences were made explicit(378). Recognising patients' preferences for information (or not), has further been found to have significant effects on patients' symptoms of anxiety and depression. Similarly, the match between patients' preferences for involvement vs. enacted involvement in decision making was found to have significant, positive effects on patients' satisfaction with care processes (379-381)and treatment anxiety (382). Therefore, it is important that we devise strategies that improve decision making to incorporate patient preferences more closely.

Whilst the UK government had initially begun to develop patient decision aids (383) in 2011 only 8 were developed, they were poorly publicised and not user friendly. The King's Fund reported that poor clinician engagement in creating patient decision aids in the UK as a main barrier to their use(384).

Overall, modern technology, if used with purpose and insight, could be deployed to empower patients and help doctors ensure the right patient receives the best treatment, based on all of the evidence and their patient's values.

The use of variation measures in healthcare has been used as a marker of healthcare quality in order to justify economic changes. However, I have shown that the currently used metrics

of variation, which do not account for population characteristics and underlying disease burden, may not provide any valuable information.

Additionally, I have shown that we may not be aligning decisions to our patients as well as we could. A more appropriate measure for quality of clinical care provided may be related to how we share our decisions.

If all treatment decisions were based around a transparent, robust and reproducible shared decision making process it would be likely to reduce unwarranted surgical rate variation. In such circumstances, any observed variation would reflect true variation in the preferences of patients treated by different general practices and surgical teams. Combining a better shared decision making process with a greater insight of regional population 'need' for policy makers may be a more appropriate governmental strategy towards regional surgical rate variation.

References

1. van Kempen MJ, Rijkers GT, Van Cauwenberge PB. The immune response in adenoids and tonsils. Int Arch Allergy Immunol. 2000 May;122(1):8–19.

2. Ashworth M, Cox K, Latinovic R, Charlton J, Gulliford M, Rowlands G. Why has antibiotic prescribing for respiratory illness declined in primary care? A longitudinal study using the General Practice Research Database. Journal of Public Health. 2004 Sep 1;26(3):268–74.

3. Lock CC, Wilson JJ, Steen NN, Eccles MM, Mason HH, Carrie SS, et al. North of England and Scotland Study of Tonsillectomy and Adeno-tonsillectomy in Children(NESSTAC): a pragmatic randomised controlled trial with a parallel non-randomised preference study. Health Technol Assess. 2010 Mar 1;14(13):1–iv.

4. Burton MJ, Glasziou PP. Tonsillectomy or adeno-tonsillectomy versus non-surgical treatment for chronic/recurrent acute tonsillitis. Cochrane Database Syst Rev. 2009 Jan 1;(1):CD001802–2.

5. Alho O-P, Koivunen P, Penna T, Teppo H, Koskela M, Luotonen J. Tonsillectomy versus watchful waiting in recurrent streptococcal pharyngitis in adults: randomised controlled trial. BMJ. 2007 May 5;334(7600):939–9.

6. Koskenkorva T, Koivunen P, Penna T, Teppo H, Alho O-P. Factors affecting quality-of-life impact of adult tonsillectomy. J Laryngol Otol. 2009 Apr 24;123(09):1010.

7. Koskenkorva T, Koivunen P, Koskela M, Niemela O, Kristo A, Alho O-P. Short-term outcomes of tonsillectomy in adult patients with recurrent pharyngitis: a randomized controlled trial. CMAJ. 2013 Apr 1.

8. Bhattacharyya NN, Kepnes LJL. Economic benefit of tonsillectomy in adults with chronic tonsillitis. Ann Otol Rhinol Laryngol. 2002 Nov 1;111(11):983–8.

9. HSCIC. Hospital Admitted Patient Care Activity. Web Master, United Kingdom.

10. Weatherly RA, Mai EF, Ruzicka DL, Chervin RD. Identification and evaluation of obstructive sleep apnea prior to adenotonsillectomy in children: a survey of practice patterns. Sleep Med. 2003 Jul;4(4):297–307.

11. Westert GP, Smits JPJM, Polder JJ, Mackenbach JP. Community income and surgical rates in the Netherlands. J Epidemiol Community Health. BMJ Publishing Group Ltd; 2003 Jul;57(7):519–22.

12. Windfuhr JP. Specified data for tonsil surgery in Germany. GMS Current Topics in Otorhinolaryngology, Head and Neck Surgery. German Medical Science; 2016;15:Doc08.

13. Brownell M. Tonsillectomy Rates for Manitoba Children: Temporal and Spatial Variations. Healthcare Management Forum. 2002 Dec;15(4):21–6.

14. Douglas-Jones P, Fagan JJ. Tonsillectomy rates in the South African private healthcare sector. South African Medical Journal. 2016 Nov 1;106(11):1134–40.

15. Boss EF, Marsteller JA, Simon AE. Outpatient tonsillectomy in children: demographic and geographic variation in the United States, 2006. J Pediatr. 2012 May;160(5):814–9.

16. Glover JA. The incidence of tonsillectomy in school children. 1938. Vol. 37, International journal of epidemiology. 2008. 11 p.

17. Bloor MJ, Venters GA, Samphier ML. Geographical variation in the incidence of operations on the tonsils and adenoids. An epidemiological and sociological investigation (Part 2). J Laryngol Otol. 1978 Oct;92(10):883–95.

18. Newton JN, Seagroatt V, Goldacre M. Geographical variation in hospital admission rates: an analysis of workload in the Oxford region, England. J Epidemiol Community Health. BMJ Publishing Group; 1994 Dec;48(6):590–5.

19. Suleman M, Clark M, Goldacre M, Burton M. Exploring the variation in paediatric tonsillectomy rates between English regions: a 5-year NHS and independent sector data analysis. Clinical Otolaryngology. Wiley Online Library; 2010;35(2):111–7.

20. Appleby J. Variations in Health Care. 2011. 1.

21. Van Den Akker EH, Hoes AW, Burton MJ, Schilder AGM. Large international differences in (adeno)tonsillectomy rates. Clin Otolaryngol. 2004 Apr;29(2):161–4.

22. Bakwin H. The tonsil-adenoidectomy enigma. J Pediatr. 1958 Mar 1;52(3):339–61.

23. Bloor M. Bishop Berkeley and the Adenotonsillectomy Enigma: An Exploration of Variation in the Social Construction of Medical Disposals. Sociology. 1976;10(1):43–61.

24. Wennberg JE, Barnes BA, Zubkoff M. Professional uncertainty and the problem of supplier-induced demand. Soc Sci Med. 1982 Jan 1;16(7):811–24.

25. Kaplan S, NASLUND M. Public, patient, and professional attitudes towards the diagnosis and treatment of enlarged prostate: a landmark national US survey. International Journal of Clinical Practice. Blackwell Publishing Ltd; 2006 Oct 1;60(10):1157–65.

26. Tarbox BB, Rockwood JK, Abernathy CM. Are modified radical mastectomies done for T1 breast cancers because of surgeon“s advice or patient”s choice? Am J Surg. 1992 Nov 1;164(5):417–412.

27. Weinstein JN. Trends: Trends And Geographic Variations In Major Surgery For Degenerative Diseases Of The Hip, Knee, And Spine. Health Aff (Millwood). 2004 Oct 7;:1–10.

28. Fisher B, Bauer M, Margolese R, Poisson R, Pilch Y, Redmond C, et al. Five-Year Results of a Randomized Clinical Trial Comparing Total Mastectomy and Segmental Mastectomy with or without Radiation in the Treatment of Breast Cancer. New England Journal of Medicine. 1985;312(11):665–73.

29. Fisher B, Redmond C, Poisson R, Margolese R, Wolmark N, Wickerham L, et al. Eight-Year Results of a Randomized Clinical Trial Comparing Total Mastectomy and Lumpectomy with or without Irradiation in the Treatment of Breast Cancer. New England Journal of Medicine. 1989;320(13):822–8.

30. Veronesi U, Saccozzi R, Del Vecchio M, Banfi A, Clemente C, De Lena M, et al. Comparing Radical Mastectomy with Quadrantectomy, Axillary Dissection, and Radiotherapy in Patients with Small Cancers of the Breast. New England Journal of Medicine. 1981;305(1):6–11.

31. Morris J, McNoe B, Elwood JM, Packer S. Breast cancer surgery in New Zealand: Consensus or variation? New Zealand Medical Journal. 1997 Dec 1;110(1038):53–6.

32. Woon YY, Chan MYP. Breast conservation surgery—the surgeon's factor. The Breast. 2005 Apr;14(2):131–5.

33. Wennberg JE, Mulley AG, Hanley D, Timothy RP, Fowler FJ, Roos NP, et al. An assessment of prostatectomy for benign urinary tract obstruction. Geographic variations and the evaluation of medical care outcomes. JAMA. 1988 May 27;259(20):3027–30.

34. Weinstein JN. Trends: Trends And Geographic Variations In Major Surgery For Degenerative Diseases Of The Hip, Knee, And Spine. Health Aff (Millwood). 2004 Oct 7.

35. Wennberg J, Gittelsohn A. Variations in medical care among small areas. Sci Am. 1982 Apr;246(4):120–34.

36. Wright JG, Hawker GA, Bombardier C, Croxford R, Dittus RS, Freund DA, et al. Physician enthusiasm as an explanation for area variation in the utilization of knee replacement surgery. Med Care. 1999 Sep;37(9):946–56.

37. Chassin MR. Explaining geographic variations. The enthusiasm hypothesis. Med Care. 1993 May;31(5 Suppl):YS37–44.

38. Krasnik A, Groenewegen PP, Pedersen PA, Scholten von P, Mooney G, Gottschau A, et al. Changing remuneration systems: effects on activity in general practice. BMJ. BMJ Publishing Group; 1990 Jun 30;300(6741):1698–701.

39. Wennberg JE, Fowler FJ. A test of consumer contribution to small area variations in health care delivery. J Maine Med Assoc. 1977 Aug;68(8):275–9.

40. Wennberg JE, Blowers L, Parker R, Gittlesohm AM. Changes in Tonsillectomy Rates Associated with Feedback and Review. Pediatrics. 1977;59(6):821–6.

41. Wennberg JE, Fowler FJ Jr. A test of consumer contribution to small a rea variations in health care delivery. J Maine Med Assox. 1977 Aug;68(8):275-9

42. Press Association. “Postcode lottery” revealed in NHS care. Guardian. Thursday 8th September 2016. https://www.theguardian.com/society/2016/sep/08/postcode-lottery-revealed-in-nhs-care

43. McCann K. Report warns of postcode lottery in NHS cancer care. The Telegraph. 27 April 2017 http://www.telegraph.co.uk/news/2016/04/26/report-warns-of-postcode-lottery-in-nhs-cancer-care/

44. Kirk A. Revealed: Shocking NHS postcode lottery for elderly care. The Telegraph. 18 September 2015. http://www.telegraph.co.uk/news/health/elder/11872521/Revealed-Shocking-NHS-postcode-lottery-for-elderly-care.html

45. McKinsey, Co. *Achieving world class productivity in the NHS 2009-10-2013-14: Detailing the size of the opportunity*; available at [Internet]. Health DO, editor. webarchive.nationalarchives.gov.uk. [cited 2017 Apr 30]. Available from: http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_116521.pdf

46. Age-standardised mortality rates by sex and region, 2001 to 2012 registrations - Office for National Statistics. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/adhocs/006541agestandardisedmortalityratesbysexandregion2001to2012registrations

47. Pope C. Resisting evidence: The study of evidence-based medicine as a contemporary social movement. Health. 2003 Jul 1;7(3):267–82.

48. McCormack L, Sheridan S, Lewis M, Boudewyns V, Melvin CL, Kistler C, et al. Communication and dissemination strategies to facilitate the use of health-related evidence. Evid Rep Technol Assess (Full Rep). 2013 Jan 1;(213):1–520.

49. Levin A. The Cochrane Collaboration. Annals of internal medicine. American College of Physicians; 2001 Aug 21;135(4):309–12.

50. Hill J, Bullock I, Alderson P. A summary of the methods that the National Clinical Guideline Centre uses to produce clinical guidelines for the National Institute for Health and

Clinical Excellence. Annals of internal medicine. American College of Physicians; 2011 Jun 7;154(11):752–7.

51. Paradise JLJ, Bluestone CDC, Bachman RZR, Colborn DKD, Bernard BSB, Taylor FHF, et al. Efficacy of tonsillectomy for recurrent throat infection in severely affected children. Results of parallel randomized and nonrandomized clinical trials. New England Journal of Medicine. 1984 Mar 15;310(11):674–83.

52. Paradise JLJ, Bluestone CDC, Colborn DKD, Bernard BSB, Rockette HEH, Kurs-Lasky MM. Tonsillectomy and adenotonsillectomy for recurrent throat infection in moderately affected children. PEDIATRICS. 2002 Jun 30;110(1 Pt 1):7–15.

53. van Staaij BK. Effectiveness of adenotonsillectomy in children with mild symptoms of throat infections or adenotonsillar hypertrophy: open, randomised controlled trial. BMJ. 2004 Sep 18;329(7467):651–0.

54. Burton MJ, Glasziou PP, Chong LY. Tonsillectomy or adenotonsillectomy versus non-surgical treatment for chronic/recurrent acute tonsillitis. The Cochrane …. 2014.

55. NATTINA - Newcastle University. Newcastle University.

56. England RCOSO. Bulletin of the Royal College of Surgeons of England. 2005. 52 p.

57. SIGN SIGN. Management of sore throat and indications for tonsillectomy. (SIGN Guideline No 117). 2010 Apr 20;1–44.

58. Clement WA, Dempster JH. Implementation by Scottish otolaryngologists of the Scottish Intercollegiate Guidelines Network document Management of Sore Throats and the Indications for Tonsillectomy: four years on. J Laryngol Otol. 2005 Oct 26;118(5):357–61.

59. Care N. NHS atlas of variation in healthcare for children and young people, 2012. 2012.

60. Birkmeyer JD, Reames BNR, McCulloch P, Carr AJ, Understanding of regional variation in the use of surgery. The Lancet; 2013 Sep 28;382(9898):1121–9.

61. Ftouh S, Morga A, Swift C. Guidelines: Management of hip fracture in adults: summary of NICE guidance. BMJ: British Medical Journal. BMJ; 2011 Jun 25;342(7812):1413–5.

62. Goodman D, Brownlee S. The Dartmouth Atlas of Health Care [Internet]. dartmouthatlas.org. [cited 2017 Apr 30]. Available from: http://www.dartmouthatlas.org/tools/downloads.aspx#surgery

63. Department of Health. *Health Profile of England 2008*

[Internet]. webarchive.nationalarchives.gov.uk. [cited 2009 Jan]. Available from: http://webarchive.nationalarchives.gov.uk/+/www.dh.gov.uk/en/Publicationsandstatistics/publications/publicationsstatistics/dh_093465

64. Rightcare. Regional diabetes data England [Internet]. england.nhs.uk. 2013 [cited 2017 Apr 30]. Available from: https://www.england.nhs.uk/statistics/statistical-work-areas/integrated-performance-measures-monitoring/diabetes-data/

65. Mainous AG, Tanner RJ, Baker R, Zayas CE, Harle CA. Prevalence of prediabetes in England from 2003 to 2011: population-based, cross-sectional study. BMJ Open. British Medical Journal Publishing Group; 2014 Jan 1;4(6):e005002–2.

66. Office of National Statistics. Adult Smoking Habits in Great Britain [Internet]. ons.gov.uk. [cited 2017 Apr 1]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/drugusealcoholandsmoking/datasets/adultsmokinghabitsingreatbritain

67. Sheikh K, Bullock C. Rise and fall of radical prostatectomy rates from 1989 to 199611The views expressed in this article do not represent the views or policies of the Centers for Medicare & Medicaid Services or the United States. Urology. 2002 Mar 1;59(3):378–82.

68. Lu-Yao G, Albertsen PC, Stanford JL, Stukel TA, Walker-Corkery ES, Barry MJ. Natural experiment examining impact of aggressive screening and treatment on prostate cancer mortality in two fixed cohorts from Seattle area and Connecticut. BMJ. British Medical Journal Publishing Group; 2002 Oct 5;325(7367):740–0.

69. Wyke S, Hewison J, Russell IT. Respiratory illness in children: what makes parents decide to consult? Br J Gen Pract. Royal College of General Practitioners; 1990 Jun;40(335):226–9.

70. Cabral C, Lucas PJ, Ingram J, Hay AD, Horwood J. “It's safer to …” parent consulting and clinician antibiotic prescribing decisions for children with respiratory tract infections: An analysis across four qualitative studies. Soc Sci Med. Elsevier Ltd; 2015 Jul;136-137(C):156–64.

71. Kai J. Parents' difficulties and information needs in coping with acute illness in preschool children: a qualitative study. BMJ. BMJ Group; 1996 Oct 19;313(7063):987–90.

72. Kai J. What worries parents when their preschool children are acutely ill, and why: a qualitative study. BMJ. BMJ Group; 1996 Oct 19;313(7063):983–6.

73. Ashford JR, Pearson NG. Who Uses the Health Services and Why? Journal of the Royal Statistical Society Series A (General). Wiley for the Royal Statistical Society; 1970 Jan 1;133(3):295–357.

74. Geertsen HR, Gray RM. Familistic Orientation and Inclination toward Adopting the Sick Role. Journal of Marriage and Family. National Council on Family Relations; 1970 Nov 1;32(4):638–46.

75. Rahe RH, Gunderson EK, Arthur RJ. Demographic and psychosocial factors in acute illness reporting. J Chronic Dis. 1970 Oct;23(4):245–55.

76. KEDWARD HB. Social class habits of consulting. Br J Prev Soc Med. BMJ Group; 1962 Jul;16(3):147–52.

77. Hawker GA, Wright JG, Coyte PC, Williams JI, Harvey B, Glazier R, et al. Determining the need for hip and knee arthroplasty: the role of clinical severity and patients' preferences. Med Care. 2001 Mar;39(3):206–16.

78. Hawker GA, Wright JG, Coyte PC, Williams JI, Harvey B, Glazier R, et al. Differences between Men and Women in the Rate of Use of Hip and Knee Arthroplasty. New England Journal of Medicine. 2000;342(14):1016–22.

79. Cutler D, Skinner J, Stern AD, Wennberg D. Physician beliefs and patient preferences: a new look at regional variation in health care spending. Cambridge, MA: National Bureau of Economic Research; 2013.

80. Chandra A, Staiger DO. Productivity spillovers in healthcare: evidence from the treatment of heart attacks. J Polit Econ. 2007;115:103–40.

81. Hudak PL, Clark JP, Hawker GA, Coyte PC, Mahomed NN, Kreder HJ, et al. "You're Perfect for the Procedure! Why Don“t You Want It?" Elderly Arthritis Patients” Unwillingness to Consider Total Joint Arthroplasty Surgery: A Qualitative Study. Medical Decision Making. 2002 Jun 1;22(3):272–8.

82. Ballantyne PJ, Gignac MAM, Hawker GA. A patient-centered perspective on surgery avoidance for hip or knee arthritis: lessons for the future. Arthritis Rheum. 2007 Feb 15;57(1):27–34.

83. Hawker GA, Guan J, Croxford R, Coyte PC, Glazier RH, Harvey BJ, et al. A prospective population-based study of the predictors of undergoing total joint arthroplasty. Arthritis Rheum. Wiley Subscription Services, Inc., A Wiley Company; 2006 Oct;54(10):3212–20.

84. Wright JG, Santaguida PL, Young N, Hawker GA, Schemitsch E, Owen JL. Patient preferences before and after total knee arthroplasty. Journal of Clinical Epidemiology. Elsevier Inc; 2010 Jul 1;63(7):774–82.

85. Ibrahim SA. Racial variations in the use of knee and hip joint replacement: an introduction and review of the most recent literature. Current Orthopaedic Practice. 2010 Mar;21(2):126–31.

86. Emejuaiwe N, Jones AC, Ibrahim SA, Kwoh CK. Disparities in joint replacement utilization: a quality of care issue. Clin Exp Rheumatol. 2007 Nov;25(6 Suppl 47):44–9.

87. Figaro MK, Russo PW, Allegrante JP. Preferences for arthritis care among urban African Americans: "I don't want to be cut". Health Psychol. American Psychological Association; 2004 May;23(3):324–9.

88. Groeneveld PW, Kwoh CK, Mor MK, Appelt CJ, Geng M, Gutierrez JC, et al. Racial differences in expectations of joint replacement surgery outcomes. Arthritis Rheum. 2008 May 15;59(5):730–7.

89. Modi CS, Veillette CJH, Gandhi R, Perruccio AV, Rampersaud YR. Factors that influence the choice to undergo surgery for shoulder and elbow conditions. Clin Orthop Relat Res. 2014 Mar;472(3):883–91.

90. Hawker GA, Wright JG, Glazier RH, Coyte PC, Harvey B, Williams JI, et al. The effect of education and income on need and willingness to undergo total joint arthroplasty. Arthritis Rheum. 2002 Dec 12;46(12):3331–9.

91. Hawker GA, Wright JG, Badley EM, Coyte PC, Toronto Arthroplasty Health Services Research Consortium. Perceptions of, and willingness to consider, total joint arthroplasty in a population-based cohort of individuals with disabling hip and knee arthritis. Arthritis Care & Research. 2004 Aug 5;51(4):635–41.

92. Hall JA, Roter DL, Katz NR. Meta-analysis of correlates of provider behavior in medical encounters. Med Care. 1988 Jul;26(7):657–75.

93. Wallen J, Waitzkin H, Stoeckle J. Physician Stereotypes about Female Health and Illness:. Women & Health. 1979 Aug 15;4(2):135–46.

94. Willems S, De Maesschalck S, Deveugele M, Derese A, De Maeseneer J. Socio-economic status of the patient and doctor–patient communication: does it make a difference? Patient Education and Counseling. 2005 Feb;56(2):139–46.

95. Verlinde E, De Laender N, De Maesschalck S, Deveugele M, Willems S. The social gradient in doctor-patient communication. International Journal for Equity in Health. BioMed Central Ltd; 2012 Mar 12;11(1):12.

96. Siminoff LA, Graham GC, Gordon NH. Cancer communication patterns and the influence of patient characteristics: Disparities in information-giving and affective behaviors. Patient Education and Counseling. 2006 Sep;62(3):355–60.

97. Eggly S, Harper FWK, Penner LA, Gleason MJ, Foster T, Albrecht TL. Patient Education and Counseling. Patient Education and Counseling. Elsevier Ireland Ltd; 2011 Jan 1;82(1):63–8.

98. Waitzkin H. Information giving in medical care. J Health Soc Behav. 1985 Jun;26(2):81–101.

99. Street RL Jr., Gordon H, Haidet P. Physicians' communication and perceptions of patients: Is it how they look, how they talk, or is it just the doctor? Social Science & Medicine [Internet]. 2007 Aug;65(3):586–98. Available from: http://www.sciencedirect.com/science/article/pii/S0277953607001645

100. Boss EF, Mehta N, Nagarajan N, Links A, Benke JR, Berger Z, et al. Shared Decision Making and Choice for Elective Surgical Care: A Systematic Review. Otolaryngology-Head and Neck Surgery. 2016 Mar;154(3):405–20.

101. Wennberg JE. Time to tackle unwarranted variations in practice. BMJ. BMJ; 2011;342.

102. Fragaszy EB, Warren-Gash C, Wang L, Copas A, Dukes O, Edmunds WJ, et al. Cohort Profile: The Flu Watch Study. Int J Epidemiol. Oxford University Press; 2017 Apr 1;46(2):e18–8.

103. Hayward AC, Fragaszy EB, Bermingham A, Wang L, Copas A, Edmunds WJ, et al. Comparative community burden and severity of seasonal and pandemic influenza: results of the Flu Watch cohort study. Lancet Respir Med. 2014 Jun;2(6):445–54.

104. Fragaszy EB, Quinlivan M, Breuer J, Craig R, Hutchings S, Kidd M, et al. Population-level susceptibility, severity and spread of pandemic influenza: design of, and initial results from, a pre-pandemic and hibernating pandemic phase study using cross-sectional data from the Health Survey for England (HSE). Vol. 3, Public Health Research. Southampton (UK): NIHR Journals Library; 2015. 24 p.

105. Marshall T. A review of tonsillectomy for recurrent throat infection. Br J Gen Pract. 2011 Jan 13;48(431):1331–5.

106. Singh S, Dolan JG, Centor RM. Optimal management of adults with pharyngitis--a multi-criteria decision analysis. BMC Medical Informatics and Decision Making. 2006;6:14.

107. Hawker JI, Smith S, Smith GE, Morbey R, Johnson AP, Fleming DM, et al. Trends in antibiotic prescribing in primary care for clinical syndromes subject to national

recommendations to reduce antibiotic resistance, UK 1995-2011: analysis of a large database of primary care consultations. J Antimicrob Chemother. 2014 Nov 12;69(12):3423–30.

108. Millett ERC, Quint JK, Smeeth L, Daniel RM, Thomas SL. Incidence of community-acquired lower respiratory tract infections and pneumonia among older adults in the United Kingdom: a population-based study. PLoS ONE. Public Library of Science; 2013;8(9):e75131–.

109. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol. 2015 Jun;44(3):827–36.

110. Carter H, Chen S, Isik L, Tyekucheva S, Velculescu VE, Kinzler KW, et al. Cancer-specific high-throughput annotation of somatic mutations: computational prediction of driver missense mutations. Cancer Res. American Association for Cancer Research; 2009 Aug 15;69(16):6660–7.

111. Herrett E, Smeeth L, Walker L, Weston C. The Myocardial Ischaemia National Audit Project (MINAP). Heart. BMJ Publishing Group Ltd; 2010 Aug 1;96(16):1264–7.

112. Smith NR, Lewis DJ, Fahy A, Eldridge S, Taylor SJ, Moore DG, et al. Individual socio-demographic factors and perceptions of the environment as determinants of inequalities in adolescent physical and psychological health: the Olympic Regeneration in East London (ORiEL) study. BMC Public Health. 2nd ed. London: Department for …; 2015 Feb 15;15(1):341.

113. Denaxas SC, George J, Herrett E, Shah AD, Kalra D, Hingorani AD, et al. Data resource profile: cardiovascular disease research using linked bespoke studies and electronic health records (CALIBER). Int J Epidemiol. 2012 Dec;41(6):1625–38.

114. Herrett E, Shah AD, Boggon R, Denaxas S, Smeeth L, van Staa T, et al. Completeness and diagnostic validity of recording acute myocardial infarction events in primary care, hospital care, disease registry, and national mortality records: cohort study. BMJ. British Medical Journal Publishing Group; 2013 May 20;346(may20 3):f2350–0.

115. CPRD [Internet]. cprd.com. [cited 2017 May 27]. Available from: https://www.cprd.com/isac/otherinfo.asp

116. Wallace P, Delaney B, Sullivan F. Unlocking the research potential of the GP electronic care record. Br J Gen Pract. 2013 Jun 1;63(611):284–5.

117. Meystre SM, Savova GK, Kipper-Schuler KC. Extracting information from textual documents in the electronic health record: a review of recent research. … Med Inform. 2008.

118. Hansell A, Hollowell J, Nichols T, McNiece R, Strachan D. Use of the General Practice Research Database (GPRD) for respiratory epidemiology: a comparison with the 4th Morbidity Survey in General Practice (MSGP4). Thorax. 1999 May;54(5):413–9.

119. Lovaasen KR. ICD-10-CM/PCS Coding: Theory and Practice, 2017 Edition. Elsevier Health Sciences; 2016. 1 p.

120. Rooney C, Griffiths C, Cook L. The implementation of ICD-10 for cause of death coding-some preliminary results from the bridge coding study. Health Statistics Quarterly. 2002.

121. McLennan D, Barnes H, Noble M, Davies J, Garratt E. The English indices of deprivation 2010. London: Department for …; 2011.

122. Statistics OON. Internet Access Households and Individuals, 2011. Statistics, National OO, editors. Statistic Bulletin. 2012 Dec 31;:1–9.

123. Martin D. Geography for the 2001 Census in England and Wales. Office for National Statistics; 2002.

124. Mathur R, Bhaskaran K, Chaturvedi N, Leon DA, vanStaa T, Grundy E, et al. Completeness and usability of ethnicity data in UK-based primary care and hospital databases. J Public Health (Oxf). 2014 Dec;36(4):684–92.

125. Bhaskaran K, Forbes HJ, Douglas I, Leon DA, Smeeth L. Representativeness and optimal use of body mass index (BMI) in the UK Clinical Practice Research Datalink (CPRD). BMJ Open. British Medical Journal Publishing Group; 2013 Sep 13;3(9):e003389.

126. Campbell J, Dedman DJ, Eaton SC, Gallagher AM, Williams TJ. Is the Cprd Gold Population Comparable to the U.k. Population?: 567. Pharmacoepidem Drug Safe. Pharmacoepidemiology and Drug Safety; 2013 Oct 1;22:280–1.

127. Copeland KT, Checkoway H, McMichael AJ, Holbrook RH. Bias due to misclassification in the estimation of relative risk. Am J Epidemiol. 1977 May;105(5):488–95.

128. Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. Br J Clin Pharmacol. Blackwell Publishing Ltd; 2010 Jan;69(1):4–14.

129. Van Staa TP, Dennison EM, Leufkens HGM, Cooper C. Epidemiology of fractures in England and Wales. Bone. 2001 Dec 19;29(6):517–22.

130. Ryan R, Majeed A. Prevalence of treated hypertension in general practice in England and Wales, 1994 to 1998. Health Statistics Quarterly. 2002.

131. Ronquist G, Rodríguez LAG, Ruigómez A, Johansson S, Wallander M-A, Frithz G, et al. Association between captopril, other antihypertensive drugs and risk of prostate cancer. Prostate. Wiley Subscription Services, Inc., A Wiley Company; 2004 Jan 1;58(1):50–6.

132. Meier CR, Napalkov PN, Wegmüller Y, Jefferson T, Jick H. Population-based study on incidence, risk factors, clinical complications and drug utilisation associated with influenza in the United Kingdom. Eur J Clin Microbiol Infect Dis. 2000 Nov;19(11):834–42.

133. Tate AR, Kalra D, Boggon R, Beloff N, Puri S, Williams T. Data quality in European primary care research databases. Report of a workshop held in London September 2013. BHI. IEEE; 2014;:85–8.

134. Nouraei SAR, Virk JS, Hudovsky A, Wathen C, Darzi A, Parsons D. Accuracy of clinician-clinical coder information handover following acute medical admissions: implication for using administrative datasets in clinical outcomes management. J Public Health (Oxf). 2016 Jun;38(2):352–62.

135. Nouraei SAR, Hudovsky A, Frampton AE, Mufti U, White NB, Wathen CG, et al. A Study of Clinical Coding Accuracy in Surgery: Implications for the Use of Administrative Big Data for Outcomes Management. Annals of surgery. 2015 Jun;261(6):1096–107.

136. Nouraei SAR, Hudovsky A, Virk JS, Chatrath P, Sandhu GS. An audit of the nature and impact of clinical coding subjectivity variability and error in otolaryngology. Clinical Otolaryngology. 2013 Dec;38(6):512–24.

137. Tai TWT, Anandarajah SS, Dhoul NN, de Lusignan SS. Variation in clinical coding lists in UK general practice: a barrier to consistent data entry? Inform Prim Care. 2007 Jan 1;15(3):143–50.

138. Marshall TT, Mohammed MAM, Lim HTH. Understanding variation for clinical governance: an illustration using the diagnosis and treatment of sore throat. Br J Gen Pract. 2002 Mar 31;52(477):277–83.

139. Koshy E, Curcin V, Bottle A, Sharland M, Saxena S. Sore throat consultations in general practice prior to tonsillectomy among eight hundred and sixty-three children in England: is this in accordance with the SIGN guidelines? Clin Otolaryngol. 2013 May 31;38(3):266–70.

140. Centre for Clinical Practice at NICE (UK). Respiratory Tract Infections - Antibiotic Prescribing: Prescribing of Antibiotics for Self-Limiting Respiratory Tract Infections in Adults and Children in Primary Care. London: National Institute for Health and Clinical Excellence (UK); 2008.

141. Heminway H. www.caliberresearch.org [Internet]. caliberresearch.org. [cited 2017 May 27]. Available from: https://www.caliberresearch.org/

142. Hawker JI, Smith S, Smith GE, Morbey R, Johnson AP, Fleming DM, et al. Trends in antibiotic prescribing in primary care for clinical syndromes subject to national recommendations to reduce antibiotic resistance, UK 1995-2011: analysis of a large database of primary care consultations. J Antimicrob Chemother. 2014 Nov 12;69(12):3423–30.

143. Petersen I, Hayward AC, on behalf of the SACAR Surveillance Subgroup. Antibacterial prescribing in primary care. Journal of Antimicrobial Chemotherapy. 2007 Aug 1;60(Supplement 1):i43–7.

144. Cornish RP, Henderson J, Boyd AW, Granell R, van Staa T, Macleod J. Validating childhood asthma in an epidemiological study using linked electronic patient records. BMJ Open. British Medical Journal Publishing Group; 2014 Apr 23;4(4):e005345.

145. Price D, Wilson AM, Chisholm A, Rigazio A, Burden A, Thomas M, et al. Predicting frequent asthma exacerbations using blood eosinophil count and other patient data routinely available in clinical practice. J Asthma Allergy. 2016;9:1–12.

146. Quint JK, Müllerova H, DiSantostefano RL, Forbes H, Eaton S, Hurst JR, et al. Validation of chronic obstructive pulmonary disease recording in the Clinical Practice Research Datalink (CPRD-GOLD). BMJ Open. British Medical Journal Publishing Group; 2014 Jul 1;4(7):e005540.

147. Micali N, Hagberg KW, Petersen I, Treasure JL. The incidence of eating disorders in the UK in 2000-2009: findings from the General Practice Research Database. BMJ Open. British Medical Journal Publishing Group; 2013 May 28;3(5):e002646.

148. Bell S, Daskalopoulou M, Rapsomaniki E, George J, Britton A, Bobak M, et al. Association between clinically recorded alcohol consumption and initial presentation of 12 cardiovascular diseases: population based cohort study using linked health records. BMJ. British Medical Journal Publishing Group; 2017 Mar 22;356:j909.

149. Booth HP, Prevost AT, Gulliford MC. Validity of smoking prevalence estimates from primary care electronic health records compared with national population survey data for England, 2007 to 2011. Pharmacoepidem Drug Safe. 2013 Dec;22(12):1357–61.

150. Wijnands JMA, van Durme CMPG, Driessen JHM, Boonen A, Klop C, Leufkens B, et al. Individuals With Type 2 Diabetes Mellitus Are at an Increased Risk of Gout But This Is Not Due to Diabetes: A Population-Based Cohort Study. Medicine (Baltimore). 2015 Aug;94(32):e1358.

151. Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. Clin Trials. 2004;1(4):368–76.

152. Marston L, Carpenter JR, Walters KR, Morris RW, Nazareth I, Petersen I. Issues in multiple imputation of missing data for large general practice clinical databases. Pharmacoepidem Drug Safe. John Wiley & Sons, Ltd; 2010 Jun;19(6):618–26.

153. Rubin DB. Inference and missing data. Biometrika. 1976 Dec 1;63(3):581–92.

154. Carpenter JR, Kenward MG. *A critique of common approaches to missing data. In: Missing data in randomised controlled trials— a practical guide*. National Institute for Health Research; 2007.

155. Vach W, Blettner M. Biased estimation of the odds ratio in case-control studies due to the use of ad hoc methods of correcting for missing values for confounding variables. Am J Epidemiol. 1991 Oct 15;134(8):895–907.

156. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. The BMJ. BMJ Publishing Group; 2009;338(jun29 1):b2393–3.

157. Pujades-Rodriguez M, Duyx B, Thomas SL, Stogiannis D, Smeeth L, Hemingway H. Associations between polymyalgia rheumatica and giant cell arteritis and 12 cardiovascular diseases. Heart. BMJ Publishing Group Ltd and British Cardiovascular Society; 2016 Mar;102(5):383–9.

158. Pujades-Rodriguez M, Duyx B, Thomas SL, Stogiannis D, Rahman A, Smeeth L, et al. Rheumatoid Arthritis and Incidence of Twelve Initial Presentations of Cardiovascular Disease: A Population Record-Linkage Cohort Study in England. Schooling CM, editor. PLoS ONE. Public Library of Science; 2016;11(3):e0151245–.

159. Crowson MG, Ryan MA, Rocke DJ, Raynor EM, Puscas L. Variation in tonsillectomy rates by health care system type. Int J Pediatr Otorhinolaryngol. 2017 Mar;94:40–4.

160. Akaike H. A New Look at the Statistical Model Identification. IEEE Transactions on Automatic Control. 1974 Jan 1;19(6):716–23.

161. McPherson KK, Wennberg JEJ, Hovind OBO, Clifford PP. Small-area variations in the use of common surgical procedures: an international comparison of New England, England, and Norway. New England Journal of Medicine. 1982 Nov 18;307(21):1310–4.

162. McPherson KK, A D, D B. Systematic Variation in Surgical Procedures and Hospital Admission Rates. London. London School of Hygiene and Tropical Medicine

. 1996 Jan 11.

163. Bevan G, Hollinghurst S, Benton P, Spark V, Sanderson H, Franklin D. Using Information on Variation in Rates of Supply to Question Professional Discretion in Public Services. Financial Accountability and Management. John Wiley and Sons, Ltd; 2004 Feb;20(1):1–17.

164. Kataria G, Saxena A, Bhagat S, Singh B, Kaur M, Kaur G. Deep Neck Space Infections: A Study of 76 Cases. Iran J Otorhinolaryngol. 2015 Jul;27(81):293–9.

165. Kilty SJ, Gaboury I. Clinical predictors of peritonsillar abscess in adults. J Otolaryngol Head Neck Surg. 2008 Apr;37(2):165–8.

166. Shoemaker M, Lampe RM, Weir MR. Peritonsillitis: abscess or cellulitis? Pediatr Infect Dis. 1986 Jul;5(4):435–9.

167. McNulty CA, Nichols T, French DP, Joshi P, Butler CC. Expectations for consultations and antibiotics for respiratory tract infection in primary care: the RTI clinical iceberg. Br J Gen Pract. 2013 Jul 1;63(612):429–36.

168. Neill SJ. Acute childhood illness at home: the parents' perspective. J Adv Nurs. 2000 Apr;31(4):821–32.

169. van Driel ML, De Sutter A, Deveugele M, Peersman W, Butler CC, De Meyere M, et al. Are sore throat patients who hope for antibiotics actually asking for pain relief? The Annals of Family Medicine. American Academy of Family Physicians; 2006 Nov;4(6):494–9.

170. Becker MH. The Health belief model and personal health behavior. Slack; 1974. 1 p.

171. Andersen RM. Revisiting the behavioral model and access to medical care: does it matter? J Health Soc Behav. 1995 Mar;36(1):1–10.

172. Shaw C. A framework for the study of coping, illness behaviour and outcomes. J Adv Nurs. Blackwell Science Ltd; 1999 May;29(5):1246–55.

173. Walley PP, Silvester KK, Steyn RR. Managing variation in demand: lessons from the UK National Health Service. J Healthc Manag. 2006 Aug 31;51(5):309–302.

174. HESonline. http://content.digital.nhs.uk/hes

175. The NHS Atlas of Variation in Healthcare. 2010. *https://fingertips.phe.org.uk/documents/Atlas_2010%20Compendium.pdf*

176. Rightcare. Measuring Shared Decision Making. 2012 Dec 18;:1–26. Available from: http://www.rightcare.nhs.uk/sdm

177. Right care Essential reading. https://www.england.nhs.uk/rightcare/what-is-nhs-rightcare/

178. Gray M. Unwarranted variation: a reading list produced by QIPP right care. Right care Essential reading. 2009 Nov 19;339(nov19 2):b4811–1.

179. Wennberg JEJ. On patient need, equity, supplier-induced demand, and the need to assess the outcome of common medical practices. Med Care. 1985 Apr 30;23(5):512–20.

180. Emanuel EJ, Emanuel LL. Four models of the physician-patient relationship. JAMA. 1992 Apr 21;267(16):2221–6.

181. Guyer B, Freedman MA, Strobino DM, Sondik EJ. Annual summary of vital statistics: trends in the health of Americans during the 20th century. PEDIATRICS. 2000 Dec 1;106(6):1307–17.

182. nice.org.uk [Internet]. [cited 2014 Feb 7]. Available from: https://www.nice.org.uk

183. Eysenbach G, Powell J, Kuss O, Sa E-R. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. JAMA. American Medical Association; 2002;287(20):2691–700.

184. Aymé S, Kole A, Groft S. Empowerment of patients: lessons from the rare diseases community. The Lancet. 2008 Jun 11;371(9629):2048–51.

185. Charles CC, Gafni AA, Whelan TT. Decision-making in the physician-patient encounter: revisiting the shared treatment decision-making model. Soc Sci Med. 1999 Aug 31;49(5):651–61.

186. Kinmonth AL, Spiegal N, Woodcock A. Interventions for providers to promote a patient-centred approach in clinical consultations (Review). 2013 Jan 1;:1–66.

187. BROWN SJ. Patient-Centered Communication. Annual review of nursing research. Springer Publishing Company; 1999.

188. Ockene JK, Ockene IS, Quirk ME, Hebert JR, Saperia GM, Luippold RS, et al. Physician training for patient-centered nutrition counseling in a lipid intervention trial. Prev Med. 1995 Nov 1;24(6):563–70.

189. Ford SS, Fallowfield LL, Lewis SS. Doctor-patient interactions in oncology. Soc Sci Med. 1996 May 31;42(11):1511–9.

190. Health DO. Liberating the NHS: No decision about me, without me. 2012 Dec 10;:1–42.

191. Elwyn G, Tsulukidze M, Edwards A, Légaré F, Newcombe R. Using a "talk" model of shared decision making to propose an observation-based measure: Observer OPTION 5 Item. Patient Education and Counseling. Elsevier Ireland Ltd; 2013 Nov;93(2):265–71.

192. Miles MB, Huberman AM. Qualitative data analysis. Sage Publications, Inc; 1984. 1 p.

193. Tversky A, Kahneman D. The framing of decisions and the psychology of choice. Science. 1981 Jan 30;211(4481):453–8.

194. Ajzen I, Heilbroner RL, Fishbein M, Thurow LC. Understanding Attitudes and Predicting Social Behaviour. Prentice Hall; 1980. 1 p.

195. Keeney RL. Keeney: andH. Raiffa - Google Scholar. Decision Analysis with Multiple Conflicting Objectives. 1976.

196. Brehm J. Decisional Problems. Science. 1977 Sep 30;197(4311):1355–6.

197. Orem DE, Taylor SG, Renpenning KM. Nursing. Mosby Inc; 2001. 1 p.

198. Feather NT. Values, valences, and course enrollment: Testing the role of personal values within an expectancy^valence framework. Journal of Educational Psychology. American Psychological Association; 1988;80(3):381–91.

199. Ottawa Decision Support Framework. 2011 Mar 29;:1–3.

200. Frederikson LG. Exploring information-exchange in consultation: the patients' view of performance and outcomes. Patient Education and Counseling. 1995 Jul;25(3):237–46.

201. Tuckett D. Meetings between experts : an approach to sharing ideas in medical consultations. London ; New York : Tavistock; 1985.

202. Frederikson LG. Development of an Integrative Model for Medical Consultation. Health Commun. 2014 Apr 28;5(3):225–37.

203. Carter WB, Inui TS, Kukull WA, Haigh VH. Outcome-based doctor-patient interaction analysis: II. Identifying effective provider and patient behavior. Med Care. 1982 Jun;20(6):550–66.

204. Engel JF, Blackwell RD, Miniard PW. Consumer Behavior. Houghton Mifflin; 1995. 1 p.

205. Levenstein JH, McCracken EC, McWhinney IR, Stewart MA, Brown JB. The patient-centred clinical method. 1. A model for the doctor-patient interaction in family medicine. Family Practice. 1986 Mar;3(1):24–30.

206. Middleton JF. The exceptional potential of the consultation revisited. J R Coll Gen Pract. Royal College of General Practitioners; 1989 Sep;39(326):383–6.

207. Lavis J, Ross S, McLeod C, Gildiner A. Measuring the impact of health research. J Health Serv Res Policy. 2003 Jul;8(3):165–70.

208. Mead N, Bower P. Patient-centredness: a conceptual framework and review of the empirical literature. Soc Sci Med. Elsevier; 2000;51(7):1087–110.

209. Engel GL. The Need for a New Medical Model: A Challenge for Biomedicine. Science. American Association for the Advancement of Science; 1977 Apr 8;196(4286):129–36.

210. Byrne PS, Long B. Doctors talking to patients: A study of the verbal behaviours of doctors in the consultation. London: Her Majesty's Stationary Office; 1976.

211. Roth A, Fonagy P. What Works for Whom?, Second Edition. Guilford Publications; 2013. 1 p.

212. BALINT M. The Doctor, His Patient, and the Illness. The American Journal of the Medical Sciences. 1957 Nov;234(4):609.

213. Committee HOCH. Patient and Public Involvement in the NHS. Third report of session; 2006.

214. Bate P, Robert G. Experience-based design: from redesigning the system around the patient to co-designing services with the patient. Qual Saf Health Care. BMJ Publishing Group Ltd; 2006 Oct;15(5):307–10.

215. Opie A. "Nobody's Asked Me for My View": Users' Empowerment by Multidisciplinary Health Teams. Qualitative Health Research. 1998.

216. McLean A. Empowerment and the psychiatric consumer/ex-patient movement in the United States: contradictions, crisis and change. Soc Sci Med. 1995 Apr;40(8):1053–71.

217. Young IM. Punishment, treatment, empowerment: three approaches to policy for pregnant addicts. Fem Stud. 1994 Mar 1;20(1):33–57.

218. chapman A. Dementia. 1993.

219. Menon ST. Toward a model of psychological health empowerment: implications for health care in multicultural communities. Nurse Educ Today. 2002 Jan 1;22(1):28–23.

220. Janis IL, Mann L. Coping with Decisional Conflict. American Scientist. 1976 Nov;64(6):657–67.

221. Gerard HB. Choice difficulty, dissonance, and the decision sequence. Journal of Personality. 1967 Jan 1;35(1):91–108.

222. Janis IL. Effects of Fear Arousal on Attitude Change: Recent Developments in Theory and Experimental Research1. Advances in Experimental Social Psychology. 1967. 59 p.

223. Janis IL, Mann L. Decision making: A psychological analysis of conflict, choice, and commitment. Free Press; 1977.

224. Payne JW, Samper A, Bettman JR, Luce MF. Boundary conditions on unconscious thought in complex decision making. Psychol Sci. 2008 Nov;19(11):1118–23.

225. Degner LF, Beaton JI. Life-Death Decisions in Health Care. eweb:70956. 1987.

226. Degner LF, Russell CA. Preferences for treatment control among adults with cancer. Res Nurs Health. 1988 Dec;11(6):367–74.

227. Sutherland HJ, Llewellyn-Thomas HA. Cancer patients: their desire for information and participation in treatment decisions. Journal of the Royal …. 1989.

228. Degner LF, Kristjanson LJ, Bowman D, Sloan JA, Carriere KC, O'Neil J, et al. Information needs and decisional preferences in women with breast cancer. JAMA. 1997 May 13;277(18):1485–92.

229. Thompson GN, Estabrooks CA, Degner LF. Clarifying the concepts in knowledge transfer: A literature review. J Adv Nurs. Blackwell Publishing Ltd; 2006 Mar 1;53(6):691–701.

230. Degner LF, Sloan JA, Venkatesh P. The Control Preferences Scale. Can J Nurs Res. 1997;29(3):21–43.

231. Bernoulli D. Exposition of a New Theory on the Measurement of Risk. Econometrica, Vol. 22, JSTOR. 1954 p.23-36

232. Ajzen I. Attitudes, personality, and behavior. Chicago, IL : Dorsey Press; 1988.

233. Neumann von J, Morgenstern O. Theory of games and economic behavior. Theory of Games and Economic Behavior. 2007.

234. Kerstholt JH, Raaijmakers JGW, Valeton JM. The effect of expectation on the identification of known and unknown persons. Applied Cognitive Psychology. John Wiley & Sons, Ltd; 1992 Jan 1;6(2):173–80.

235. Emanuel LL, Emanuel EJ, Stoeckle JD, Hummel LR, Barry MJ. Advance directives. Stability of patients' treatment choices. Arch Intern Med. 1994 Jan 24;154(2):209–17.

236. Weissman JS, Haas JS, Fowler FJ, Gatsonis C, Massagli MP, Seage GR, et al. The stability of preferences for life-sustaining care among persons with AIDS in the Boston Health Study. Medical Decision Making. 1999 Jan;19(1):16–26.

237. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. Springer Netherlands; 2010 May;19(4):539–49.

238. Services UDOHAH. Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Final Guidance. 2009. … Reported Outcome Measures: Use …; 2009.

239. Smith SCS, Lamping DLD, Banerjee SS, Harwood RR, Foley BB, Smith PP, et al. Measurement of health-related quality of life for people with dementia: development of a new instrument (DEMQOL) and an evaluation of current methodology. Health Technol Assess. 2005 Mar 1;9(10):1–iv.

240. Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, Perrin E, et al. Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res. 2002 May;11(3):193–205.

241. Elwyn G, Frosch D, Rollnick S. Dual equipoise shared decision making: definitions for decision and behaviour support interventions. Implementation Sci. 2009;4.

242. Guimond P, Bunn H, O'Connor AM, Jacobsen MJ, Tait VK, Drake ER, et al. Validation of a tool to assess health practitioners' decision support and communication skills. Patient Education and Counseling. 2003 Jun 30;50(3):235–45.

243. Singh S, Butow P, Charles M, Tattersall MHN. Shared decision making in oncology: assessing oncologist behaviour in consultations in which adjuvant therapy is considered after primary surgical treatment. Health Expectations. 2010 Jun 23;:no–no.

244. Stacey D, Taljaard M, Drake ER, O'Connor AM. Audit and feedback using the brief Decision Support Analysis Tool (DSAT-10) to evaluate nurse–standardized patient encounters. Patient Education and Counseling. 2008 Dec;73(3):519–25.

245. Kriston L, Scholl I, lzel LH, Simon D, Loh A, rter MH. Patient Education and Counseling. Patient Education and Counseling [Internet]. Elsevier Ireland Ltd; 2010 Jul 1;80(1):94–9. Available from: http://www.sciencedirect.com/science/article/pii/S0738399109004509

246. Gagnon MM, Hibert RR, Dubé MM, Dubois M-FM. Development and validation of an instrument measuring individual empowerment in relation to personal health care: the Health Care Empowerment Questionnaire (HCEQ). Am J Health Promot. 2006 Jun 30;20(6):429–35.

247. Smoliner A, Hantikainen V, Mayer H, Ponocny-Seliger E, Them C. [Patients' preferences and experience regarding participation in nursing care decisions in acute hospitals--an analysis of conformity of preferences and experience, and factors influencing different types of decision making]. Pflege. 2009 Dec;22(6):411–9.

248. Melbourne E, Sinclair K, Durand M-A, Légaré F, Elwyn G. Developing a dyadic OPTION scale to measure perceptions of shared decision making. Patient Education and Counseling. 2010 Feb 1;78(2):177–83.

249. Saba GW. Shared Decision Making and the Experience of Partnership in Primary Care. The Annals of Family Medicine. 2006 Jan 1;4(1):54–62.

250. Silvester J, Patterson F, Koczwara A, Ferguson E. “Trust me...”: psychological and behavioral predictors of perceived physician empathy. J Appl Psychol. 2007 Mar 1;92(2):519–27.

251. Entwistle VA, Watt IS. Patient involvement in treatment decision-making: the case for a broader conceptual framework. Patient Education and Counseling. 2006 Nov 1;63(3):268–78.

252. Scholl I, Loon MK-V, Sepucha K, Elwyn G, Légaré F, Härter M, et al. Measurement of shared decision making – a review ofinstruments. Theriogenology. unknown; 2011;105(4):313–24.

253. Elwyn G, Hutchings H, Edwards A, Rapport F, Wensing M, Cheung W-Y, et al. The OPTION scale: measuring the extent that clinicians involve patients in decision-making tasks. Health Expect. 2005 Mar 1;8(1):34–42.

254. Couët N, Desroches S, Robitaille H, Vaillancourt H, Leblanc A, Turcotte S, et al. Assessments of the extent to which health-care providers involve patients in decision making: a systematic review of studies using the OPTION instrument. Health Expect. 2013 Mar 4;:–.

255. Melbourne E, Roberts S, Durand M-A, Newcombe R, Légaré F, Elwyn G. Dyadic OPTION: Measuring perceptions of shared decision-making in practice. Patient Education and Counseling. Elsevier Ireland Ltd; 2011 Apr;83(1):55–7.

256. Lerman CEC, Brody DSD, Caputo GCG, Smith DGD, Lazaro CGC, Wolfson HGH. Patients' Perceived Involvement in Care Scale: relationship to attitudes about illness and medical care. J Gen Intern Med. 1990 Jan 1;5(1):29–33.

257. Martin LR, Robin Di Matteo M, Lepper HS. Facilitation of Patient Involvement in Care: Development and Validation of a Scale. Behavioral Medicine. 2001 Jan;27(3):111–20.

258. Giersdorf N, Loh A, Bieber C, Caspari C, Deinzer A, Doering T, et al. [Development and validation of assessment instruments for shared decision making]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 2004 Sep 30;47(10):969–76.

259. Gafni A, Charles C, McMaster University. Centre for Health Economics and Policy Analysis, Whelan T. Shared Decision-making in the Medical Encounter : what Does it Mean?, Or, It Takes at Least Two to Tango. Hamilton, Ont. : Centre for Health Economics and Policy Analysis, McMaster University; 1994. 12 p.

260. Simon D, Schorr G, Wirtz M, Vodermaier A, Caspari C, Neuner B, et al. Development and first validation of the shared decision-making questionnaire (SDM-Q). Patient Education and Counseling. 2006 Nov 1;63(3):319–27.

261. Glass KE, Wills CE, Holloman C, Olson J, Hechmer C, Miller CK, et al. Shared decision making and other variables as correlates of satisfaction with health care decisions in a United States national survey. Patient Education and Counseling. 2012 Jul;88(1):100–5.

262. Scholl I, Kriston L, Dirmaier J, Buchholz A, Härter M. Development and psychometric properties of the Shared Decision Making Questionnaire--physician version (SDM-Q-Doc). Patient Education and Counseling. 2012 Jul 31;88(2):284–90.

263. O'Connor AM. Decisional Conflict Scale. 1993 Nov 27;:1–1.

264. O'Connor AM. Validation of a decisional conflict scale. Medical Decision Making. 1995 Jan 1;15(1):25–30.

265. Thompson-Leduc P, Turcotte S, Labrecque M, Légaré F. Prevalence of clinically significant decisional conflict: an analysis of five studies on decision-making in primary care. BMJ Open. British Medical Journal Publishing Group; 2016 Jun 28;6(6):e011490.

266. Sun Q. Predicting downstream effects of high decisional conflict: Meta-analyses of the decisional conflict scale. University of Ottawa (Canada); 2005.

267. Knops AM, Goossens A, Ubbink DT, Legemate DA, Stalpers LJ, Bossuyt PM. Interpreting patient decisional conflict scores: behavior and emotions in decisions about treatment. Med Decis Making. 2013 Jan;33(1):78–84.

268. O'Connor AM. User Manual – Decisional Conflict Scale. 2014 May 8;:1–16.

269. Dolan JG. A Method for Evaluating Health Care Providers' Decision Making. Medical Decision Making. 1999.

270. Stacey D, Bennett CL, Barry MJ, Col NF, Eden KB, Holmes-Rovner M, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev. 2011 Jan 1;(10):CD001431–1.

271. Wennberg JE. Dealing with medical practice variations: a proposal for action. Health Aff (Millwood). 1984;3(2):6–32.

272. Modi CS, Veillette CJH, Gandhi R, Perruccio AV, Rampersaud YR. Factors That Influence the Choice to Undergo Surgery for Shoulder and Elbow Conditions. Clin Orthop Relat Res. 2013 Nov 2;472(3):883–91.

273. Johnson RM. Beyond Conjoint Measurement: a Method of Pairwise Trade-Off Analysis. ACR North American Advances. 1976;NA-03.

274. Saaty TL. Optimum positions for m airports. Naval Res Logist Quart. Wiley Subscription Services, Inc., A Wiley Company; 1972;19(1):101–9.

275. Saaty RW. The analytic hierarchy process-what it is and how it is used. Mathematical Modelling. 1987 Jan 1;9(3-5):161–76.

276. Liberatore MJ, Nydick RL. The analytic hierarchy process in medical and health care decision making: A literature review. European Journal of Operational Research. 2008 Aug;189(1):194–207.

277. Schmidt K, Aumann I, Hollander I, Damm K, Schulenburg von der J-MG. Applying the Analytic Hierarchy Process in healthcare research - A systematic literature review and evaluation of reporting. BMC Medical Informatics and Decision Making. BioMed Central; 2015;15(1):112.

278. Marsh K, Lanitis T, Neasham D, Orfanos P, Caro J. Assessing the value of healthcare interventions using multi-criteria decision analysis: a review of the literature. Pharmacoeconomics. Springer International Publishing; 2014 Apr;32(4):345–65.

279. Sculpher MJ, Dwyer N, Browning J, Horsley S, Cullimore J. A survey of women's preferences regarding alternative surgical treatments for menorrhagia. Health Expectations. Blackwell Science Ltd; 1998 Nov;1(2):96–105.

280. Sculpher M. Economic Evaluation in Health Care. 2001.

281. Crow R, Gage H, Hampson S, Hart J, Kimber A, Storey L, et al. The measurement of satisfaction with healthcare: implications for practice from a systematic review of the literature. Health Technol Assess. 2002;6(32):1–244.

282. Amos T, Daniel K. Prospect Theory: An Analysis of Decision under Risk. Econometrica. 1979;47(2):263.

283. Torrance GW, Feeny D. Utilities and Quality-Adjusted Life Years. Int J Technol Assess Health Care. Cambridge University Press; 1989 Oct 1;5(4):559–75.

284. Duru G, Auray JP, Béresniak A, Lamure M, Paine A, Nicoloyannis N. Limitations of the methods used for calculating quality-adjusted life-year values. Pharmacoeconomics. 2002;20(7):463–73.

285. Farquhar PH, Keller LR. Preference intensity measurement. Annals of Operations Research. 2nd ed. Baltzer Science Publishers, Baarn/Kluwer Academic Publishers; 1989 Dec;19(1):205–17.

286. Torrance GW. Measurement of health state utilities for economic appraisal. J Health Econ. 1986 Mar;5(1):1–30.

287. Au Eong KG, Chan EW, Luo N, Wong SH, Tan NWH, Lim TH, et al. Validity of EuroQOL-5D, time trade-off, and standard gamble for age-related macular degeneration in the Singapore population. Eye (Lond). 2012 Mar;26(3):379–88.

288. Badia X, Monserrat S, Roset M, Herdman M. Feasibility, validity and test-retest reliability of scaling methods for health states: the visual analogue scale and the time trade-off. Qual Life Res. 1999 Jun;8(4):303–10.

289. Lissitz RW, Green SB. Effect of the number of scale points on reliability: A Monte Carlo approach. J Appl Psychol. American Psychological Association; 1975 Feb 1;60(1):10–3.

290. Lozano LM, García-Cueto E, Muñiz J. Effect of the number of response categories on the reliability and validity of rating scales. Methodology. Hogrefe & Huber Publishers; 2008 Dec 1;4(2):73–9.

291. Miller GA. The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychological Review [Internet]. 1956;63(2):81–97.

292. Rao VR. Beyond Conjoint Analysis: Advances in Preference Measurement. In: Applied Conjoint Analysis. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013. pp. 363–82.

293. Netzer O, Toubia O, Bradlow ET, Dahan E, Evgeniou T, Feinberg FM, et al. Beyond conjoint analysis: Advances in preference measurement. Marketing Letters. Springer; 2008 Dec 1;19(3/4):337–54.

294. Mulye R. An empirical comparison of three variants of the AHP and two variants of conjoint analysis. J Behav Dec Making. 1998 Jan 1;11(4):263–80.

295. Bech M, Kjaer T, Lauridsen J. Does the number of choice sets matter? Results from a web survey applying a discrete choice experiment. Health Econ. John Wiley & Sons, Ltd; 2011 Mar;20(3):273–86.

296. Nunnally. Psychometric Theory 3E. Tata McGraw-Hill Education; 2010. 1 p.

297. Scholl A, Manthey L, Helm R, Steiner M. Solving multiattribute design problems with analytic hierarchy process and conjoint analysis: An empirical comparison. European Journal of Operational Research. 2005 Aug;164(3):760–77.

298. Ekman P. An Argument for Basic Emotions. Cognition and Emotion. 1992 May 1;6(3-4):169–200.

299. Llewellyn-Thomas H, Sutherland HJ. The measurement of patients' values in medicine. Medical decision Making. 1982; 2(4):449-62

300. Schoemaker PJH. The Expected Utility Model: Its Variants, Purposes, Evidence and Limitations. Journal of Economic Literature. American Economic Association; 1982 Jun 1;20(2):529–63.

301. Meißner M, Scholz SW, Decker R. AHP versus ACA - An empirical comparison. In: Preisach C, Burkhardt H, Schmidt-Thieme L, Decker R, editors. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008. pp. 447–54.

302. Helm R, Scholl A, Manthey L, Steiner M. Measuring customer preferences in new product development: comparing compositional and decompositional methods. International Journal of Product Development. 2004 Jan 1;1(1):12–29.

303. Whitaker R. Validation examples of the Analytic Hierarchy Process and Analytic Network Process. Mathematical and Computer Modelling. 2007 Oct;46(7-8):840–59.

304. Saaty TL. How to Make a Decision: The Analytic Hierarchy Process. Interfaces. INFORMS; 1994 Nov 1;24(6):19–43.

305. Critical Appraisal Skills Programme (2017) [Internet]. docs.wixstatic.com. [cited 2017 Jun 10]. Available from: http://docs.wixstatic.com/ugd/dded87_afbfc99848f64537a53826e1f5b30b5c.pdf

306. Critical Appraisal Skills Programme (2017) [Internet]. docs.wixstatic.com. [cited 2017 Jun 10]. Available from: http://docs.wixstatic.com/ugd/dded87_861b48c94b654b82a84250ca684d9186.pdf

307. Critical Appraisal Skills Programme (2017) [Internet]. docs.wixstatic.com. [cited 2017 Jun 10]. Available from: http://docs.wixstatic.com/ugd/dded87_5ad0ece77a3f4fc9bcd3665a7d1fa91f.pdf

308. Critical Appraisal Skills Programme (2017) [Internet]. docs.wixstatic.com. [cited 2017 Jun 10]. Available from: http://docs.wixstatic.com/ugd/dded87_7e983a320087439e94533f4697aa109c.pdf

309. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Academic Press; 1977. 1 p.

310. bob P, ball C. Oxford Centre for Evidence-based Medicine - Levels of Evidence (March 2009) - CEBM. 2009 Jun 11.

311. Hunt SD, Richard D Sparkman SJS, Wilcox JB. The Pretest in Survey Research: Issues and Preliminary Findings. Journal of Marketing Research. American Marketing Association; 1982 May 1;19(2):269–73.

312. Wennberg DED, Wennberg JEJ. Addressing variations: is there hope for the future? Health Aff (Millwood). 2003 Jun 30;Suppl Web Exclusives:W3–W7.

313. O'Connor AM, Tugwell P, Wells GA, Elmslie T, Jolly E, Hollingworth G, et al. A decision aid for women considering hormone therapy after menopause: decision support framework and evaluation. Patient Education and Counseling. 1998 Mar;33(3):267–79.

314. Dolan JG, Frisina S. Randomized Controlled Trial of a Patient Decision Aid for Colorectal Cancer Screening. Medical Decision Making. 2002.

315. Goel V, Sawka CA, Thiel EC, Gort EH, O'Connor AM. Randomized trial of a patient decision aid for choice of surgical treatment for breast cancer. Medical Decision Making. 2001;21(1):1–6.

316. Nascimento MCV, Carvalho ACPLF. A graph clustering algorithm based on a clustering coefficient for?weighted graphs. Journal of the Brazilian Computer Society. 2010 Dec 21;17(1):19–29.

317. Mooi E, Sarstedt M. Cluster Analysis. In: A Concise Guide to Market Research. Berlin, Heidelberg: Springer Berlin Heidelberg; 2010. pp. 237–84.

318. Dolan JG. Patient priorities in colorectal cancer screening decisions. Blackwell Science Ltd; 2005. pp. 334–44.

319. Valls A, Torra V. Using classification as an aggregation tool in MCDM. Fuzzy Sets and Systems. 2000;115(1):159–68.

320. Baldé A, Légaré F, Labrecque M. Assessment of needs of men for decision support on male sterilization. Patient Education and Counseling. 2006 Nov;63(3):301–7.

321. Hageman MGJS, Döring A-C, Spit SA, Guitton TG, Ring D. Assessment of Decisional Conflict about the Treatment of Trigger Finger, Comparing Patients and Physicians. Arch Bone Jt Surg. 2016 Oct;4(4):353–8.

322. Lam WWT, Chan M, Or A, Kwong A, Suen D, Fielding R. Reducing treatment decision conflict difficulties in breast cancer surgery: a randomized controlled trial. Journal of Clinical Oncology. 2013 Aug 10;31(23):2879–85.

323. Dales RE, O'Connor A, Hebert P, Sullivan K, McKim D, Llewellyn-Thomas H. Intubation and Mechanical Ventilation for COPD: Development of an Instrument To Elicit Patient Preferences. Chest. American College of Chest Physicians; 1999 Sep 1;116(3):792–800.

324. Clark HD, O'Connor AM, Graham ID, Wells GA. What factors are associated with a woman's decision to take hormone replacement therapy? Evaluated in the context of a decision aid. Health Expect. 2003 Jun;6(2):110–7.

325. Flynn KE, Weinfurt KP, Seils DM, Lin L, Burnett CB, Schulman KA, et al. Decisional conflict among patients who accept or decline participation in phase I oncology studies. J Empir Res Hum Res Ethics. 2008 Sep;3(3):69–77.

326. Gruppen LD, Margolin J, Wisdom K, Grum CM. Outcome bias and cognitive dissonance in evaluating treatment decisions. Acad Med. 1994 Oct;69(10 Suppl):S57–9.

327. Festinger L, Carlsmith JM. Cognitive consequences of forced compliance. The Journal of Abnormal and Social Psychology. American Psychological Association; 1959 Mar 1;58(2):203–10.

328. Gooberman-Hill R, Sansom A, Sanders CM, Dieppe PA, Horwood J, Learmonth ID, et al. Unstated factors in orthopaedic decision-making: a qualitative study. BMC Musculoskeletal Disorders. 2nd ed. BioMed Central; 2010 Sep 17;11(1):213.

329. Sonntag U, Brink A, Renneberg B, Braun V, Heintze C. GPs' attitudes, objectives and barriers in counselling for obesity--a qualitative study. Eur J Gen Pract. 8 ed. 2012 Mar;18(1):9–14.

330. Say RE, Thomson R. The Importance Of Patient Preferences In Treatment Decisions: Challenges For Doctors. BMJ: British Medical Journal. BMJ; 2003 Sep 6;327(7414):542–5.

331. McNeil BJ, Pauker SG, Sox HC Jr., Tversky A. On the Elicitation of Preferences for Alternative Therapies. N Engl J Med. 1982 May 27;306(21):1259–62.

332. Kirklin D. Framing, truth telling and the problem with non-directive counselling. J Med Ethics. Institute of Medical Ethics; 2007 Jan;33(1):58–62.

333. Chewning B, Bylund CL, Shah B, Arora NK, Gueguen JA, Makoul G. Patient preferences for shared decisions: a systematic review. Patient Education and Counseling. Elsevier Ireland Ltd; 2012 Jan;86(1):9–18.

334. Xu Y, Levy BT, Daly JM, Bergus GR, Dunkelberg JC. Comparison of patient preferences for fecal immunochemical test or colonoscopy using the analytic hierarchy process. BMC Health Serv Res. BioMed Central; 2015 Apr 23;15(1):175–9.

335. Danner M, Vennedey V, Hiligsmann M, Fauser S, Gross C, Stock S. Comparing Analytic Hierarchy Process and Discrete Choice Experiment to Elicit Patient Preferences for Treatment Characteristics in Age-Related Macular Degeneration. Value in Health. 2017 May 31.

336. Danner M, Vennedey V, Hiligsmann M, Fauser S, Gross C, Stock S. How Well Can Analytic Hierarchy Process be Used to Elicit Individual Preferences? Insights from a Survey in

Patients Suffering from Age-Related Macular Degeneration. Patient. Springer International Publishing; 2016 Oct;9(5):481–92.

337. Forman EH, Gass SI. The Analytic Hierarchy Process - An Exposition. Operations Research. INFORMS; 2001;49(4):469–86.

338. Mühlbacher AC, Bethge S, Kaczynski A, Juhnke C. Objective Criteria in the Medicinal Therapy for Type II Diabetes: An Analysis of the Patients' Perspective with Analytic Hierarchy Process and Best-Worst Scaling. Gesundheitswesen. © Georg Thieme Verlag KG; 2016 May 1;78(5):326–36.

339. Kuruoglu E, Guldal D, Mevsim V, Gunvar T. Which family physician should I choose? The analytic hierarchy process approach for ranking of criteria in the selection of a family physician. BMC Medical Informatics and Decision Making. BioMed Central; 2015;15(1):63.

340. IJzerman MJ, van Til JA, Bridges JFP. A comparison of analytic hierarchy process and conjoint analysis methods in assessing treatment alternatives for stroke rehabilitation. Patient. Springer International Publishing; 2012;5(1):45–56.

341. IJzerman MJ, van Til JA, Snoek GJ. Comparison of two multi-criteria decision techniques for eliciting treatment preferences in people with neurological disorders. Patient. 2008 Dec 1;1(4):265–72.

342. Holder RD. Some Comments on the Analytic Hierarchy Process. The Journal of the Operational Research Society. Operational Research Society; 1990 Nov 1;41(11):1073–6.

343. Dong Y, Xu Y, Li H, Dai M. A comparative study of the numerical scales and the prioritization methods in AHP. European Journal of Operational Research. 2008;186(1):229–42.

344. Ji P, Jiang R. Scale transitivity in the AHP. J Oper Res Soc. 2003 Jul 24;54(8):896–905.

345. Ishizaka A, Balkenborg D, Kaplan T. Influence of aggregation and measurement scale on ranking a compromise alternative in AHP. J Oper Res Soc. Palgrave Macmillan UK; 2010;62(4):700–10.

346. Finan JS, Hurley WJ. Transitive calibration of the AHP verbal scale. European Journal of Operational Research. 1999;112(2):367–72.

347. Lootsma FA. Conflict resolution via pairwise comparison of concessions. European Journal of Operational Research. 1989 May 5;40(1):109–16.

348. Harker PT, Vargas LG. THEORY OF RATIO SCALE ESTIMATION: SAATY'S ANALYTIC HIERARCHY PROCESS. Manage Sci. 1987 Nov 1;33(1):1383–403.

349. Cohen J1-1. Statistical power analysis for the behavioral sciences. New York, Academic Press; 1969.

350. Banks MHM, Beresford SAS, Morrell DCD, Waller JJJ, Watkins CJC. Factors influencing demand for primary medical care in women aged 20-44 years: a preliminary report. Int J Epidemiol. 1975 Aug 31;4(3):189–95.

351. Kvestad EE, Kvaerner KJK, Røysamb EE, Tambs KK, Harris JRJ, Magnus PP. Heritability of recurrent tonsillitis. Arch Otolaryngol Head Neck Surg. 2005 Apr 30;131(5):383–7.

352. Falagas ME, Vardakas KZ, Mourtzoukou EG. Sex differences in the incidence and severity of respiratory tract infections. Respiratory Medicine. 2008 Apr 1;102(4):627.

353. Chen Y, Williams E, Kirk M. Risk Factors for Acute Respiratory Infection in the Australian Community. Schildgen O, editor. PLoS ONE. 2014 Jul 17;9(7):e101440–7.

354. Lunn JE, Knowelden J, Roe JW. Patterns of respiratory illness in Sheffield junior schoolchildren. A follow-up study. J Epidemiol Community Health. BMJ Publishing Group Ltd; 1970 Nov 1;24(4):223–8.

355. Lunn JE. Respiratory measurements of 3,556 Sheffield schoolchildren. Br J Prev Soc Med. 1965.

356. Wiseman CE, Baker R. Exploration of population and practice characteristics explaining differences between practices in the proportion of hospital admissions that are emergencies. BMC Fam Pract. Third. 2014 May 21;15(1):101.

357. Scantlebury R, Rowlands G, Durbaba S, Schofield P, Sidhu K, Ashworth M. Socioeconomic deprivation and accident and emergency attendances: cross-sectional

analysis of general practices in England. Br J Gen Pract. British Journal of General Practice; 2015 Oct;65(639):e649–54.

358. Katzburg JR, Yano EM, Washington DL, Farmer MM, Yee EFT, Fu S, et al. Combining women's preferences and expert advice to design a tailored smoking cessation program. Substance Use and Misuse. 2009 Nov 23;44(14):2114–27.

359. Sanelli PC, Gold RL, Greenberg ED, Reichman MB, Ugorec I, Segal AZ, et al. Work-in-progress toward incorporating patients' preferences in practice guidelines for imaging aneurysmal subarachnoid hemorrhage. Acad Radiol. 2009 May;16(5):535–40.

360. Haynes B, Haines A. Barriers and Bridges to Evidence-Based Clinical Practice. In: Getting Research Findings into Practice: Second Edition. Oxford, UK: Blackwell Publishing Ltd; 2008. pp. 115–22.

361. Ellrodt G, Cook DJ, Lee J, Cho M, Hunt D, Weingarten S. Evidence-based disease management. JAMA. 1997 Nov 26;278(20):1687–92.

362. Guyatt GH, Haynes RB, Jaeschke RZ, Cook DJ, Green L, Naylor CD, et al. Users' Guides to the Medical Literature. JAMA. American Medical Association; 2000 Sep 13;284(10):1290–6.

363. Turk DC, Salovey P. Reasoning, Inference, and Judgment in Clinical Psychology. 1988. 1 p.

364. Daniel K. Judgment under uncertainty. Kahneman D, Slovic P, Tversky A, editors. Cambridge: Cambridge University Press; 2009.

365. Davis DA, Thomson MA, Oxman AD, Haynes RB. Changing physician performance. A systematic review of the effect of continuing medical education strategies. JAMA. 1995 Sep 6;274(9):700–5.

366. Gerteis M, Picker/Commonwealth Program for Patient-Centered Care. Through the patient's eyes : understanding and promoting patient-centered care. San Francisco : Jossey-Bass; 1993.

367. Eraker SA, Kirscht JP, Becker MH. Understanding and improving patient compliance. Annals of internal medicine. 1984 Jan 1;100(2):258–68.

368. Klevorick AK, Raiffa H. Decision Analysis: Introductory Lectures on Choices Under Uncertainty. The Journal of Finance. 1969 Dec;24(5):1000.

369. Pauker SG, McNeil BJ. Impact of patient preferences on the selection of therapy. J Chronic Dis. 1981 Jan 1;34(2-3):77–86.

370. Osborn R, Moulds D, Schneider EC, Doty MM, Squires D, Sarnak DO. Primary Care Physicians In Ten Countries Report Challenges Caring For Patients With Complex Health Needs. Health Aff (Millwood). Project HOPE - The People-to-People Health Foundation, Inc; 2015 Dec;34(12):2104–12.

371. Brennan PF, Strombom I. Improving Health Care by Understanding Patient Preferences: The Role of Computer Technology. Journal of the American Medical Informatics Association. 1998 May 1;5(3):257–62.

372. Barry MJ, Fowler FJ, Mulley AG, Henderson JV, Wennberg JE. Patient reactions to a program designed to facilitate patient participation in treatment decisions for benign prostatic hyperplasia. Med Care. 1995 Aug;33(8):771–82.

373. Lenert LA, Soetikno RM. Automated computer interviews to elicit utilities: potential applications in the treatment of deep venous thrombosis. Journal of the American Medical Informatics Association. American Medical Informatics Association; 1997 Jan;4(1):49–56.

374. Goldstein MK, Clarke AE, Michelson D, Garber AM, Bergen MR, Lenert LA. Developing and Testing a Multimedia Presentation of a Health-state Description. Medical Decision Making. 1994 Oct 1;14(4):336–44.

375. Gustafson DH, Hawkins RP, Boberg EW, Bricker E, Pingree S, Chan CL. The use and impact of a computer-based support system for people living with AIDS and HIV infection. Proc Annu Symp Comput Appl Med Care. American Medical Informatics Association; 1994;:604–8.

376. Kaplan S. The future of patient input into medical decision making. Quality Review Bulletin. 1992 Jun 1;18(6):182.

377. Auerbach SM, Martelli MF, Mercuri LG. Anxiety, information, interpersonal impacts, and adjustment to a stressful health care situation. J Pers Soc Psychol. American Psychological Association; 1983;44(6):1284–96.

378. Martelli MF, Auerbach SM, Alexander J, Mercuri LG. Stress management in the health care setting: matching interventions with patient coping styles. J Consult Clin Psychol. 1987 Apr;55(2):201–7.

379. Harvey RM, Kazis L, Lee AFS. Decision-making preference and opportunity in VA ambulatory care patients: Association with patient satisfaction. Res Nurs Health. John Wiley & Sons, Inc; 1999 Feb;22(1):39–48.

380. Heyland DK, Cook DJ, Rocker GM, Dodek PM, Kutsogiannis DJ, Peters S, et al. Decision-making in the ICU: perspectives of the substitute decision-maker. Intensive Care Med. Springer-Verlag; 2003 Jan;29(1):75–82.

381. Keating NL, Guadagnoli E, Landrum MB, Borbas C, Weeks JC. Treatment Decision Making in Early-Stage Breast Cancer: Should Surgeons Match Patients' Desired Level of Involvement? Journal of Clinical Oncology. American Society of Clinical Oncology; 2016 Sep 21;20(6):1473–9.

382. Gattellari M, Butow PN, Tattersall MH. Sharing decisions in cancer care. Soc Sci Med. Elsevier; 2001;52(12):1865–78.

383. Health DO. Online tools aid patient health decisions [Internet]. gov.uk. [cited 2017 Jun 11]. Available from: https://www.gov.uk/government/news/online-tools-aid-patient-health-decisions

384. Nancy J Devlin. Getting the most out of PROMs: health outcomes and NHS decision-making, The King's Fund, March 2010. 2010 Mar 5;:1–92.

385. Andreou N, Hadjisymeou S, Panesar J. Does tonsillectomy improve quality of life in adults? A systematic literature review. J Laryngol Otol. Cambridge University Press; 2013 Mar 1;127(04):332–8.

386. Kara CO, Tümkaya F, Ardic N, Topuz B. Does tonsillectomy reduce the risk of being a habitual or severe snorer? Eur Arch Otorhinolaryngol. 3rd ed. 2008 Oct;265(10):1263–8.

387. Subramaniam V, Kumar P. Impact of tonsillectomy with or without adenoidectomy on the acoustic parameters of the voice: a comparative study. Arch Otolaryngol Head Neck Surg. American Medical Association; 2009 Oct;135(10):966–9.

388. Al-Abbasi AM. Tonsillectomy for the treatment of halitosis. Niger J Med. 2009 Jul;18(3):295–8.

389. Stathas T, Mallis A, Naxakis S, Mastronikolis NS, Gkiogkis G, Xenoudakis D, et al. Taste function evaluation after tonsillectomy: a prospective study of 60 patients. Eur Arch Otorhinolaryngol. Springer-Verlag; 2010 Sep;267(9):1403–7.

390. Kirstilä V, Häkkinen P, Jentsch H, Vilja P, Tenovuo J. Longitudinal analysis of the association of human salivary antimicrobial agents with caries increment and cariogenic micro-organisms: A two-year cohort study. Journal of Dental Research. 1998 Dec 1;77(1):73–80.

391. group W. The World Health Organization Quality of Life Assessment (WHOQOL): development and general psychometric properties. Soc Sci Med. 1998 May 31;46(12):1569–85.

392. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. Springer-Verlag; 1951;16(3):297–334.

393. Streiner DL, Norman GR. Health measurement scales. Oxford University Press, USA; 1995. 1 p.

394. Lohr KN. Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res. Springer; 2002;11(3):193–205.

395. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. Control Clin Trials. 1991 Aug;12(4 Suppl):142S–158S.

396. Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. Med Care. 1990 Jul;28(7):632–42.

397. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. J Chronic Dis. 1987;40(2):171–8.

398. Rodenburg-Vandenbussche S, Pieterse AH, Kroonenberg PM, Scholl I, van der Weijden T, Luyten GPM, et al. Dutch Translation and Psychometric Testing of the 9-Item Shared Decision Making Questionnaire (SDM-Q-9) and Shared Decision Making Questionnaire-Physician Version (SDM-Q-Doc) in Primary and Secondary Care. PLoS ONE. Public Library of Science; 2015;10(7):e0132158.

399. Barr PJ, Thompson R, Walsh T, Grande SW, Ozanne EM, Elwyn G. The psychometric properties of CollaboRATE: a fast and frugal patient-reported measure of the shared decision-making process. J Med Internet Res. JMIR Publications Inc., Toronto, Canada; 2014 Jan 3;16(1):e2.

400. Katapodi MC, Munro ML, Pierce PF, Williams RA. Psychometric testing of the decisional conflict scale: genetic testing hereditary breast and ovarian cancer. Nursing Research. NIH Public Access; 2011 Nov;60(6):368–77.

401. Sim JA, Shin JS, Park SM, Chang YJ, Shin A, Noh DY, et al. Association between information provision and decisional conflict in cancer patients. Ann Oncol. 2015 Sep;26(9):1974–80.

402. Mancini J, Santin G, Chabal F, Julian-Reynier C. Cross-Cultural Validation of the Decisional Conflict Scale in a Sample of French Patients. Qual Life Res. Springer; 2006 Aug 1;15(6):1063–8.

Appendix A – Fluwatch representativeness

Table 35 Flu Watch representativeness to national cohort. This table shows the sociodemographic breakdown of FluWatch cohort compared to the national sample

	National	Nov 2006 to Mar 2007 Season 1	Nov 2007 to Mar 2008 Season 2	Nov 2008 to Mar 2009 Season 3	May 2009 to Sep 2009 Season 4	Oct 2009 to Feb 2010 Season 5	Nov 2010 to Mar 2011 Season 6
Sociodemographic variable	ONS data (% of total UK pop)	N(%)	N(%)	N(%)	N(%)	N(%)	N(%)
Age group							
0 to 4 years	6	38 (6)	42 (5)	37 (5)	36 (5)	179 (5)	45 (5)
5 to 15	11	87 (14)	110 (14)	99 (14)	109 (14)	501 (14)	131 (15)
16 to 44	42	151 (25)	258 (33)	172 (24)	192 (24)	848 (24)	206 (23)
45 to 64	25	203 (34)	272 (35)	267 (37)	293 (37)	1225 (34)	344 (38)
65+	16	123 (20)	97 (12)	154 (21)	167 (21)	799 (22)	175 (19)
Gender							
Male	49	281 (47)	366 (47)	340 (47)	377 (47)	1740 (49)	455 (51)
Female	51	321 (53)	413 (53)	389 (53)	420 (53)	1812 (51)	446 (50)
Region							
North	28	99 (16)	89 (11)	100 (14)	106 (13)	320 (9)	115 (13)
West Midlands	11	42 (7)	96 (12)	46 (6)	53 (7)	179 (5)	53 (6)

East & East Midlands	20	122 (20)	120 (15)	124 (17)	118 (15)	1456 (41)	321 (36)
London	15	28 (5)	77 (10)	26 (4)	28 (4)	270 (7)	65 (7)
South East	16	100 (17)	117 (15)	107 (15)	155 (19)	319 (9)	110 (12)
South West	10	211 (35)	280 (36)	326 (45)	337 (42)	1008 (28)	237 (26)
Vaccination status							
Vaccinated		115 (19)	130 (17)	169 (23)	0 (0)	157 (4)	186 (21)
Unvaccinated		462 (77)	632 (81)	527 (72)	797 (100)	3159 (89)	715 (79)
Unknown		25 (4)	17 (2)	33 (5)	0 (0)	236 (7)	0 (0)
Index of Multiple Deprivation quintile							
1 (most deprived)	20	37 (6)	39 (5)	28 (4)	18 (2)	98 (3)	29 (3)
2	20	88 (15)	126 (16)	91 (13)	62 (8)	310 (9)	82 (9)
3	20	164 (27)	235 (30)	238 (33)	146 (18)	915 (26)	221 (25)
4	20	162 (27)	250 (32)	187 (26)	146 (18)	938 (26)	280 (31)
5 (least deprived)	20	151 (25)	129 (17)	185 (25)	425 (53)	1291 (56)	289 (32)
Ethnicity White	75	557 (98)	733 (95)	666 (99)	730 (99)	3306 (98)	846 (98)
Non-White	25	5 (2)	3 (5)	6 (1)	7 (1)	78 (2)	19 (2)

Appendix B – Flu Watch Data management

See Figure B 1 Data Management

File **A1** (*Dailydata.dta*) contained daily data on illnesses, including sore throat occurrence, severity and additional symptoms. Each row denoted a day of illness. This file contained 37,489 rows of data describing events from 3270 patients. These data were captured from daily diaries in the first three seasons and online questionnaires in the last 2 seasons. File **B1** (*weeklydata.dta*) was the weekly data for all participants, whether they had reported an illness during that week or not. Each row related to a patient-week. Each row held information of whether the participant had a respiratory illness, but not about the specific symptoms. In addition, each contained patient information, such as socioeconomic status, ethnicity, age, co-morbidities and smoking status derived from the baseline survey.

File A1 was collapsed by illness, so that each row contained an illness episode, the new file was called **A2** (*illnessdata.dta*). File **B1** was collapsed by patient ID so each row related to one patient only. It held information regarding the dates they entered and left the study as well as dates of respiratory illnesses. This new file was labelled **B2** (*cohortdata.dta*).

Files A2 and B2 were merged by patient ID to create file **C1**. This file contained the cohort information of File B2 and the collapsed daily symptom illness data of file A2, with each row representing a period of illness, or if the patient never reported an illness in the daily data file a period of disease free observation. 83 patients were removed who had daily illness data but no matching cohort information. There were 38 illnesses reported in the weekly data for which there was no daily data, so sore throat status could not be ascertained. These weeks were not included in the denominator. Please refer to Figure B 1 Data Management for visual representation of the data management process.

For patients who reported illnesses I expanded a row prior to their first reported illness and following on from their finally reported illness. This created rows for inception and terminal periods where the patient was at risk but without any illness.

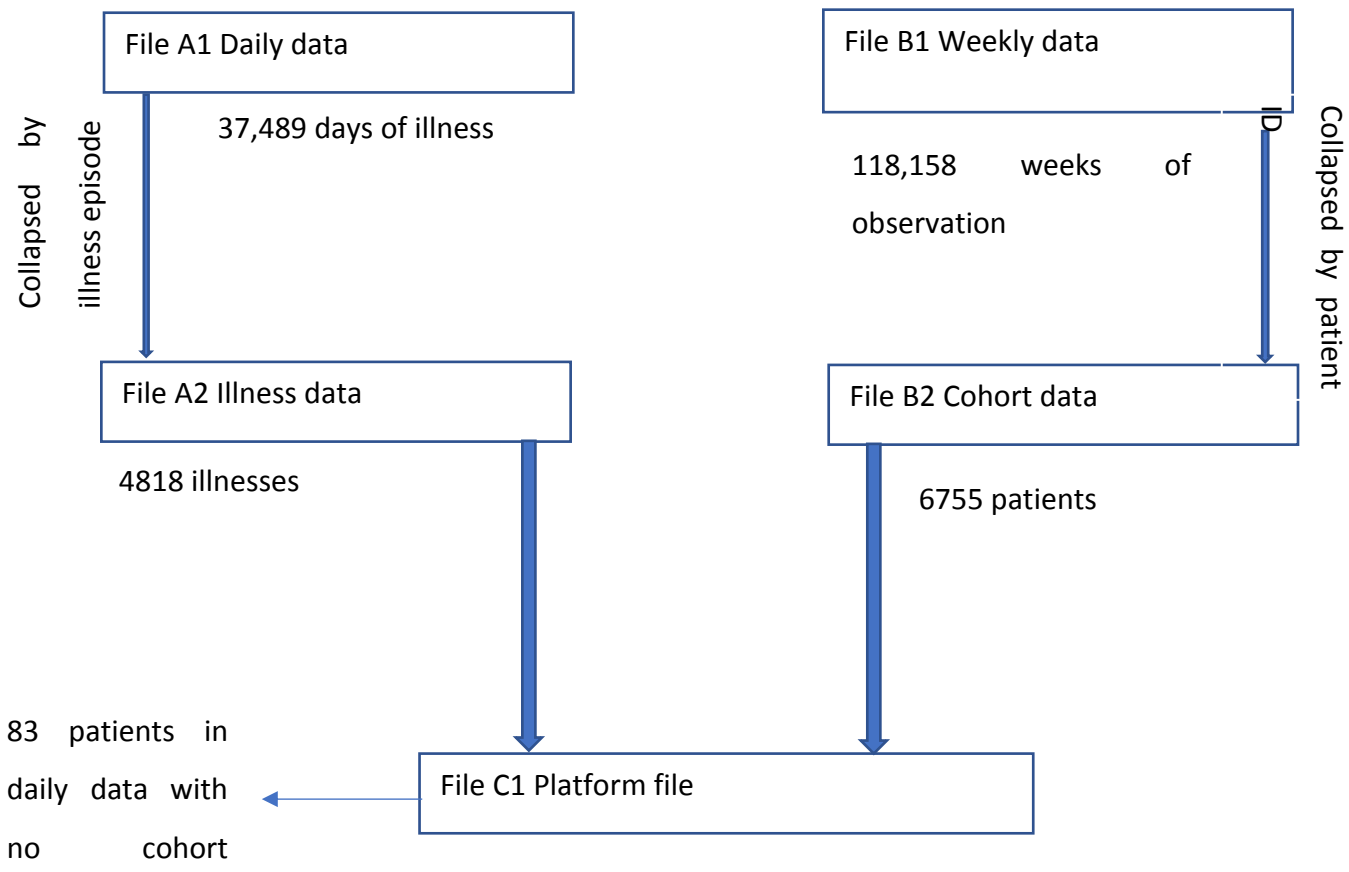


Figure B 1 Data Management

Appendix C – CALIBER Data Management

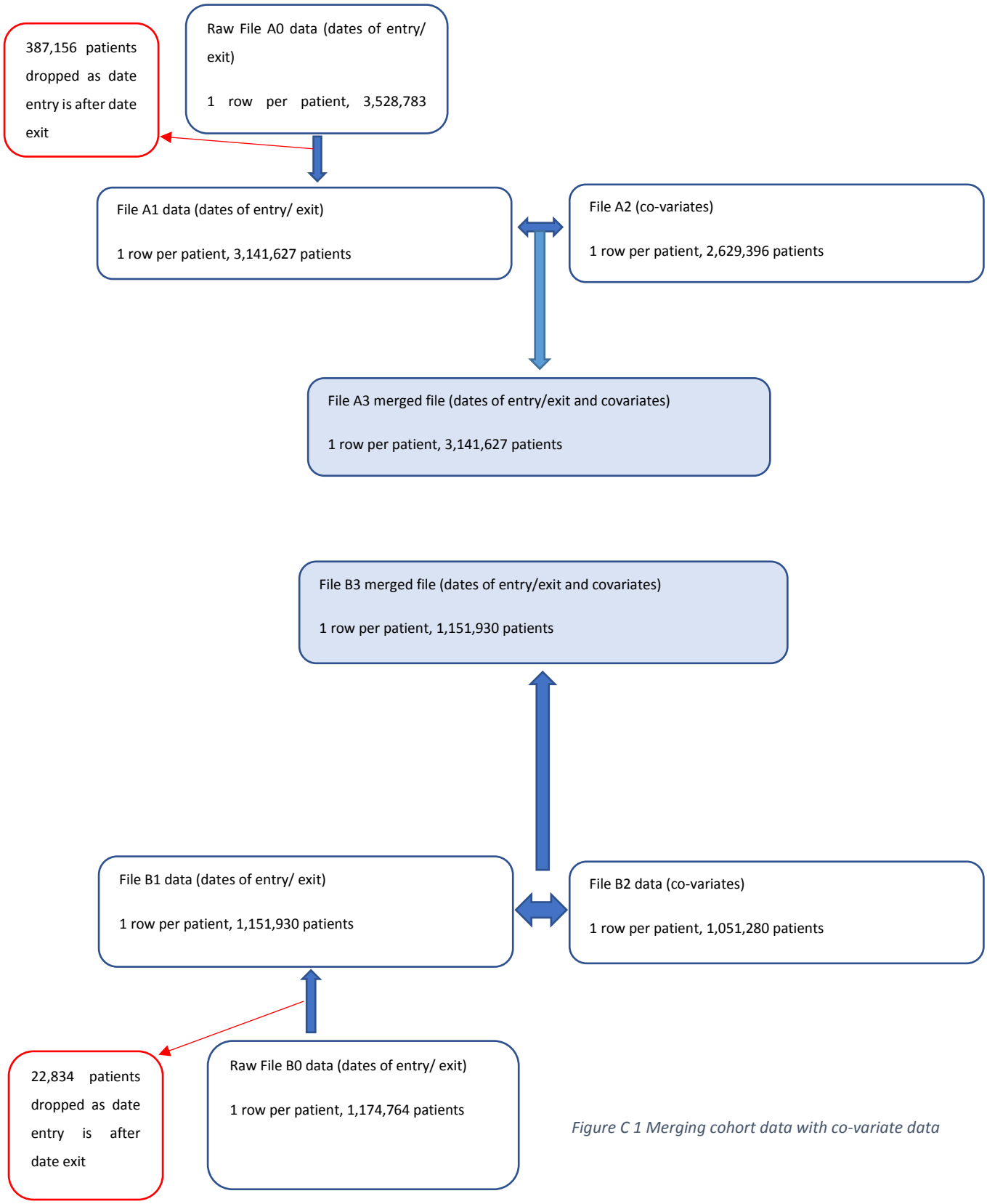


Figure C 1 Merging cohort data with co-variate data

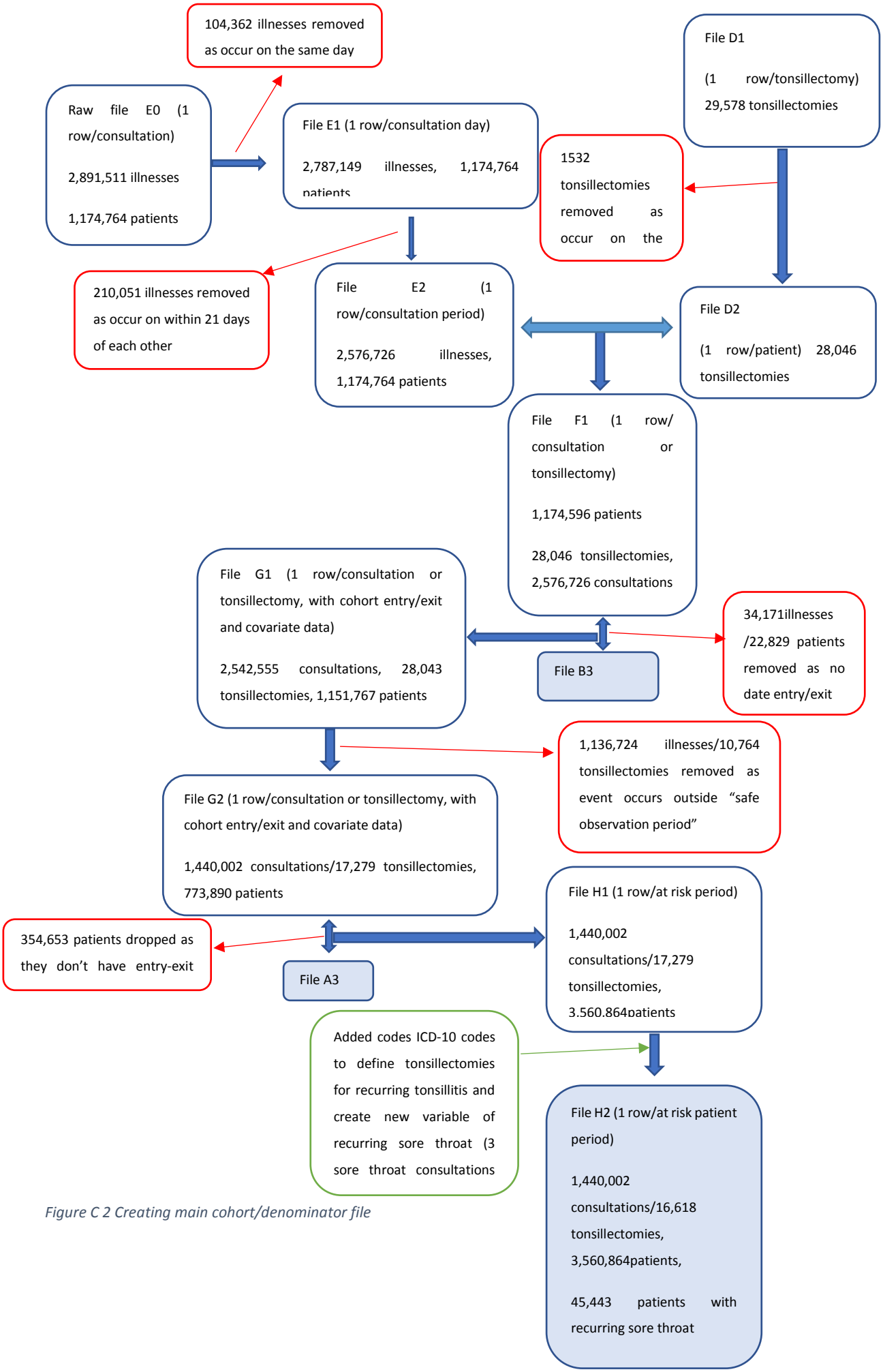


Figure C 2 Creating main cohort/denominator file

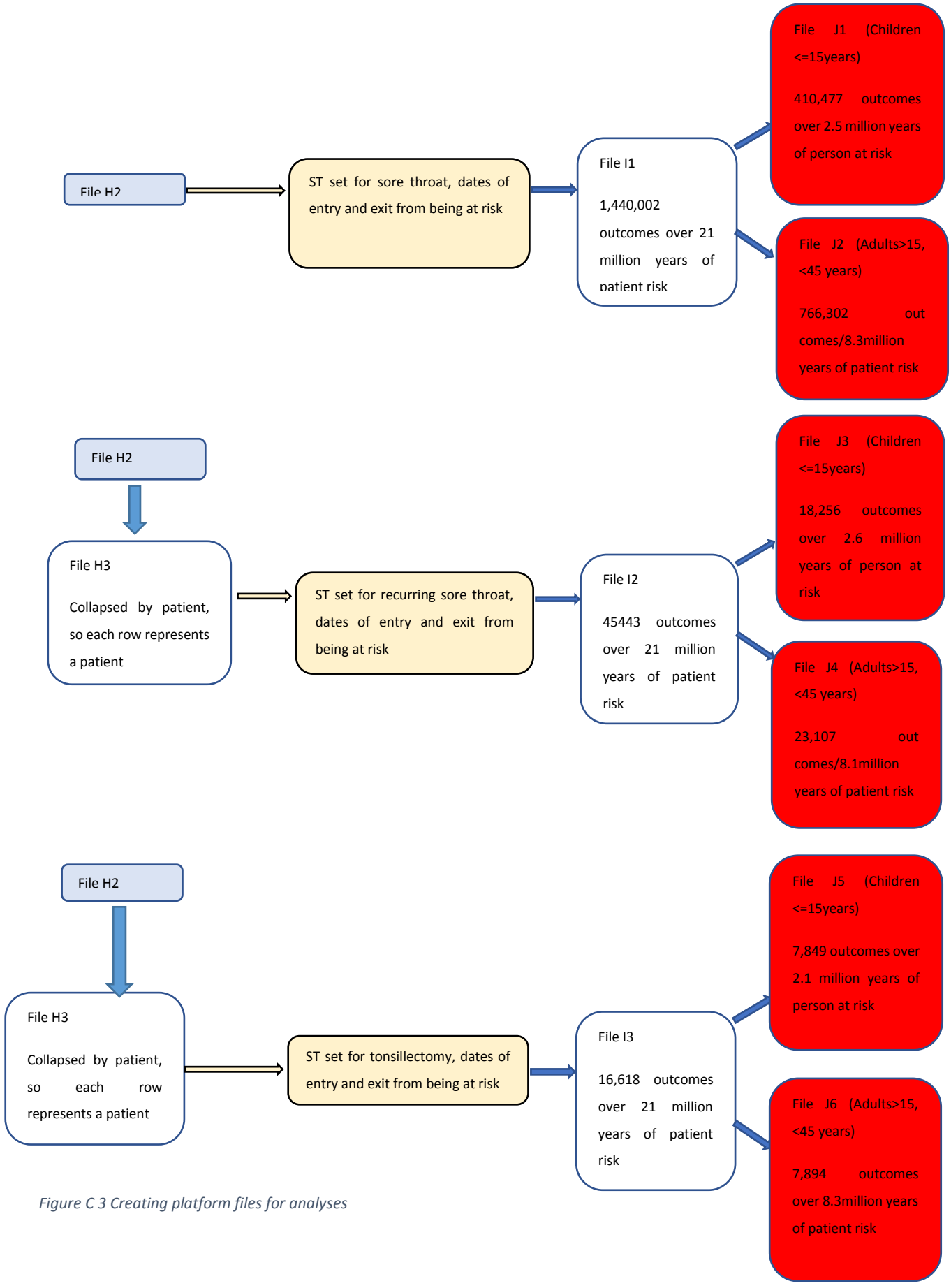


Figure C 3 Creating platform files for analyses

Appendix D – CALIBER tables

Table 36 Univariable and Multivariable multilevel Poisson analyses of sore throat consultation in children from CALIBER database

Analysis of sore throat in children							
Characteristic	Rate date			Univariable analysis		Multivariable analysis	
	Sore throat consultations	Person years	Incidence/1000 person years (95%CI)	Incidence Rate Ratio(95%CI)	P	Adjusted IRR	P
Gender							
Male	219,640	1,499,044	147 (146-148)	1		1	
Female	256,593	1,492,573	172 (171-173)	1.17(1.16-1.18)	<0.001	1.24(1.23-1.26)	<0.001
Age category							
0-4 years	117,443	469,760	250 (248-252)	1		1	
5-15 years	358,790	2,521,857	142 (141-143)	0.44(0.43-0.45)	<0.001	0.55(0.55-0.56)	<0.001
Ethnic origin							
White British	189,873	1,065,931	178 (177-179)	1		1	<0.001

Indian	2,603	1,5891	164 (154-174)	0.92(0.87-0.98)	0.87(0.82-0.93)
Black African	1,794	14,970	120 (111-128)	0.67(0.63-0.72)	0.63(0.59-0.68)
Black Caribbean	1,352	11,237	120 (111-131)	0.68(0.62-0.74)	0.68(0.62-0.74)
Black other	1,232	9,833	125 (115-136)	0.70(0.65-0.76)	0.69(0.63-0.75)
Bangladeshi	870	4,408	197 (175-222)	1.11(0.98-1.25)	1.01(0.90-1.13)
Pakistani	2,888	15,023	192 (182-203)	1.08(1.02-1.14)	0.96(0.91-1.02)
Other Asian	1,333	8,431	158 (145-172)	0.89(0.82-0.96)	0.82(0.75-0.89)
Chinese	442	3,502	126 (109-146)	0.71(0.61-0.82)	0.72(0.62-0.82)
Mixed	2,713	21,417	127 (120-134)	0.71(0.67-0.75)	0.67(0.63-0.71)

Other	3,965	26,344	151 (143-158)	0.84(0.80-0.89)		0.85(0.81-0.89)	
Unknown	75,857	489,367	155 (153-157)	0.87(0.86-0.88)	<0.001	0.92(0.90-0.93)	
Social Deprivation							
Least deprived	82,468	494,541	167 (165-169)	1		1	
2nd least deprived	93,744	570,205	164 (163-166)	0.99(0.97-1.00)		1.02(1.00-1.04)	
3rd least deprived	88,094	536,367	164 (162-166)	0.98(0.97-1.00)		1.02(1.00-1.04)	
4th least deprived	101,536	651,511	156 (154-157)	0.93(0.92-0.95)		0.97(0.95-0.97)	
Most deprived	108,186	723,119	150 (148-151)	0.89(0.88-0.91)	<0.001	0.94(0.92-0.96)	<0.001
Respiratory illness							
Absent	347,558	2,449,502	142 (141-143)	1		1	

Present	128,675	54,2115	237 (235-240)	1.67(1.66-1.69)	<0.001	1.79(1.76-1.81)	<0.001
Obstructive sleep apnoea							
Absent	474,861	2,986,318	159(158-160)	1	<0.001	1	<0.001
Present	1,363	5,211	262(239-286)	1.64(1.51-1.80)		1.79(1.76-1.81)	
Obesity							
Not coded	463,220	2,954,201	157 (156-158)	1		1	
Obese	13,013	37,416	348 (338-358)	2.22(2.15-2.29)	<0.001	2.00(1.93-2.08)	<0.001
HIV status							
HIV negative	18204	2100000	159 (158-160)	1		1	
HIV positive	52	4346	214 (196-234)	1.35(1.23-1.47)	<0.001	1.30(1.17-1.45)	<0.001
Eating disorder							
Absent	472,792	2,981,027	159 (158-159)	1		1	

Eating disorder	3,441	10,590	325 (307-343)	2.05(1.94-2.17)	<0.001	1.88(1.76-2.01)	<0.001
Practice region							
North East	2,936	22,103	133 (125-141)	1		1	
North West	82,343	477,419	172 (171-174)	1.30(1.22-1.38)		1.23(1.15-1.32)	
Yorkshire & The Humber	21,675	137,260	158 (154-161)	1.19(1.12-1.27)		1.14(1.06-1.23)	
East Midlands	20,633	117,059	176 (172-180)	1.33(1.25-1.41)		1.32(1.22-1.42)	
West Midlands	69,561	377,557	184 (182-187)	1.39(1.31-1.47)		1.37(1.27-1.47)	
East of England	65,962	403,374	164 (161-166)	1.23(1.16-1.31)		1.24(1.16-1.34)	
South West	52,192	356,662	146 (144-148)	1.10(1.04-1.17)		1.07(0.99-1.15)	
South Central	47,103	360,151	131 (129-132)	0.98(0.93-1.05)		1.07(0.99-1.15)	

London	59,959	399,297	151 (148-152)	1.13(1.07-1.20)		1.17(1.09-1.26)	
South East Coast	53,869	340,734	158 (156-160)	1.19(1.12-1.27)	<0.001	1.26(1.017-1.35)	<0.001

Table 37 Univariable and Multivariable multilevel Poisson analyses of recurring sore throat (as measured by 3 GP sore throat consultations in 1 year) in children from CALIBER database

Analysis of Rec. sore throat in children							
Characteristic	Rate date			Univariable analysis		Multivariable analysis	
	Patients with rec. sore throat	Person years	Incidence/1000 person years (95%CI)	Incidence Rate Ratio(95%CI)	P	Adjusted IRR	P
Gender							

Male	8303	1,467,697	62.2(60.9-63.4)	1	<0.001	1	<0.001
Female	9953	1,457,505	81.6(80.0-83.1)	1.31(1.28-1.35)		1.40(1.36-1.45)	
Age category							
0-4 years	4871	466,907	104.3(101.4-109.3)	1	<0.001	1	<0.001
5-15 years	16,141	2,458,295	65.7(64.7-66.7)	0.63(0.61-0.65)		0.64(0.62-0.67)	
Ethnic origin							
White British	9,715	1,036,277	93.7(91.9-95.6)	1	<0.001	1	
Indian	104	15,562	66.8(55.1-81.0)	0.70 (0.58-0.86)		0.67(0.55- 0.82)	
Black African	66	14,785	44.6(35.1-56.8)	0.44 (0.35-0.57)		0.42(0.33- 0.54)	
Black Caribbean	50	11,040	45.3(34.3-59.8)	0.49 (0.37-0.65)		0.50(0.38- 0.67)	
Black other	56	9,680	57.8(44.5-75.2)	0.62 (0.47-0.82)		0.61(0.47- 0.81)	

Bangladeshi	49	4,255	115.2(87.0-152.8)	1.14 (0.85-1.52)		1.12(0.84-1.48)	
Pakistani	147	14,619	100.1(85.6-118.2)	1.04 (0.88-1.23)		0.90(0.76- 1.06)	
Other Asian	64	7,534	77.4(60.6-98.8)	0.77 (0.60-0.98)		0.70(0.54- 0.90)	
Chinese	13	8,273	37.7(21.9-65.0)	0.35 (0.20-0.64)		0.38(0.21- 0.68)	
Mixed	105	3,446	49.8(41.1-60.3)	0.48 (0.39-0.58)		0.45(0.37- 0.55)	
Other	190	21,098	73.8(64.0-85.1)	0.79 (0.79-0.91)		0.79(0.68- 0.92)	
Unknown	3,542	25,737	74.2(71.8-76.6)	0.78 (0.75-0.81)		0.83(0.80- 0.87)	
Social Deprivation							
Least deprived	3671	483,133	75.9(73.6-78.5)	1	<0.001	1	<0.001
2nd least deprived	4166	557,436	74.7(72.5-77.0)	0.98(0.94-1.03)		1.03(0.97-1.08)	

3rd least deprived	4099	523,682	78.3(75.9-80.7)	1.03(0.99-1.08)		1.06(0.01-1.12)	
4th least deprived	4397	637,016	69.0(67.0-71.1)	0.91(0.99-1.08)		0.93(0.88-0.98)	
Most deprived	4586	708,304	64.7(62.9-66.6)	0.85(0.82-0.89)		0.89(0.84-.094)	
Respiratory illness							
Absent	14669	2,404,546	61.0(60.0-62.0)	1	<0.001	1	<0.001
Present	6343	520,656	121.8(118.9-124.9)	2.00(1.94-2.06)		2.03(1.95-2.10)	
Obstructive sleep apnoea							
Absent	20,926	2,920,327	71.7(70.7-72.6)	1		1	<0.001
Present	86	4,875	176.4(142.8-217.9)	2.46(1.99-3.04)	<0.001	2.10(1.69-2.63)	
Obesity							
Not coded	20,289	2,889,942	70.2(69.2-71.2)	1	<0.001	1	<0.001
Obese	723	35,259	205.1(190.6-220.6)	2.92(2.71-3.15)		2.31(2.12-2.53)	
HIV status							

HIV negative	20,955	2,919,423	71.8(70.8-72.8)	1	0.03	-	-
HIV positive	57	5,959	95.7(73.8-124.0)	1.33(1.03-1.73)		-	
Eating disorder							
Absent	20,823	2,915,165	71.4(70.5-72.4)	1	<0.001	1	<0.001
Eating disorder	189	10,037	188.3(163.3-217.2)	2.64(2.28-3.04)		2.30(1.95-2.70)	
Practice region							
North East	105	21,729	48.2(39.9-58.5)	1	<0.001	1	<0.001
North West	3,699	465,631	79.4(76.9-82.0)	1.64 (1.33-2.01)		1.41(1.11- 1.78)	
Yorkshire & The Humber	1,026	134,118	76.5(72.0-81.3)	1.56 (1.26-1.94)		1.42(1.13-1.79)	
East Midlands	983	113,987	86.2(81.0-91.8)	1.86 (1.50-2.31)		1.65(1.31-2.08)	
West Midlands	3,174	367,061	86.5(83.5-89.5)	1.80 (1.46-2.22)		1.68(1.33- 2.13)	

East of England	2,935	393,646	74.6(71.9-77.3)	1.55	(1.26-1.91)	1.49(1.17- 1.88)
South West	2,241	349,903	64.0(61.4-66.8)	1.25	(1.01-1.54)	1.11(0.88- 1.41)
South Central	1,884	354,759	53.1(50.8-55.6)	1.05	(0.85-1.30)	1.01(0.79- 1.28)
London	2,552	391,228	65.2(62.7-67.8)	1.29	(1.05-1.59)	1.29(1.01- 1.63)
South East Coast	2,413	333,141	72.4(69.6-75.4)	1.48	(1.20-1.83)	1.49(1.18- 1.89)

Table 38Univariable and Multivariable multilevel Poisson analyses of tonsillectomy in children from CALIBER database

Analysis of tonsillectomy in children			
Characteristic	Rate date	Univariable analysis	Multivariable analysis

	Tonsillectomies	Person years	Incidence/1000 person years (95%CI)	Incidence Rate Ratio(95%CI)	P	Adjusted IRR	P
Gender							
Male	3077	1,337,792	2.3(2.2-2.4)	1	0.15	1	<0.001
Female	2847	1,284,218	2.2(2.1-2.3)	0.96(0.92-1.01)		1.10(1.05-1.16)	
Age category							
0-4 years	208	338,542	0.6(0.5-0.7)	1	<0.001	1	<0.001
5-15 years	5716	2,283,468	2.5(2.4-2.7)	4.07(3.55-4.68)		5.34(4.64-6.15)	
Ethnic origin							
White British	4171	935,992	4.5(4.3-4.6)	1	<0.001	1	<0.001
Indian	53	14,903	3.6(2.7-4.6)	0.80(0.61-1.04)		0.86(0.66-1.13)	
Black African	41	14,478	2.8(2.1-3.8)	0.64(0.47-0.86)		0.71(0.52-0.97)	

Black Caribbean	25	10,518	2.4(1.6-3.5)	0.53(0.36-0.79)	0.50(0.34-0.74)
Black other	37	9,240	4.0(2.9-5.5)	0.90(0.65-1.24)	0.86(0.62-1.20)
Bangladeshi	12	4,186	2.9(1.6-5.0)	0.64(0.37-1.12)	0.61(0.35-1.06)
Pakistani	69	14,273	4.8(3.8-6.1)	1.08(0.86-1.37)	1.01(0.80-1.28)
Other Asian	29	8,164	3.6(2.5-5.1)	0.80(0.56-1.14)	0.94(0.65-1.34)
Chinese	6	3,292	1.8(0.8-4.0)	0.41(0.19-0.90)	0.47(0.21-1.04)
Mixed	72	20,228	3.6(2.9-4.5)	0.80(0.64-1.00)	0.92(0.73-1.16)
Other	97	24,653	3.9(3.2-4.8)	0.88(0.72-1.08)	0.89(0.73-1.08)
Unknown	1312	442,812	3.0(2.8-3.1)	0.66(0.63-0.71)	0.62(0.58-0.66)

Social Deprivation							
Least deprived	1197	436,424	2.7(2.6-2.9)	1	<0.001	1	<0.001
2nd least deprived	1323	498,228	2.7(2.5-2.8)	0.97(0.90-1.05)		0.98(0.90-1.06)	
3rd least deprived	1112	467,952	2.4(2.2-2.5)	0.87(0.80-0.94)		0.88(0.81-0.96)	
4th least deprived	1145	569,515	2.0(1.9-2.1)	0.73(0.68-0.79)		0.76(0.70-0.82)	
Most deprived	1109	635,545	1.7(1.6-1.9)	0.64(0.59-0.69)		0.68(0.62-0.74)	
Respiratory illness							
Absent	4290	2,202,847	1.9(1.9-2.0)	1	<0.001	1	<0.001
Present	1634	419,163	3.9(3.7-4.1)	2.00(1.90-2.12)		1.50(1.42-1.59)	
Obstructive sleep apnoea							
Absent	5725	2,617,329	2.2(2.1-2.2)	1	<0.001	1	<0.001

Present	199	4681	42.5(37.9-47.6)	19.44(17.30-21.84)		10.99(9.7-12.40)	
Obesity							
Not coded	5777	2,598,203	2.2(2.2-2.3)	1	<0.001	1	<0.001
Obese	147	23,807	2.2(2.2-2.3)	2.78(2.37-3.26)		1.96(1.66-2.30)	
HIV status							
HIV negative	5905	2,616,617	2.3(2.2-2.3)	1	0.05	-	
HIV positive	19	5,393	3.5(2.3-5.5)	1.56(1.00-2.43)		-	
Eating disorder							
Absent	5888	2,615,131	2.3(2.2-2.3)	1	<0.001	1	0.01
Eating disorder	36	6,879	5.2(3.8-7.2)	2.32(1.1.69-3.20)		1.54(1.13-2.12)	
Practice region							
North East	24	19,480	1.2(0.8-1.8)	1	<0.001	1	<0.001

North West	1065	406,526	2.6(2.5-2.8)	2.13(1.42-3.18)	1.92(1.30-2.84)
Yorkshire & The Humber	289	119,881	2.4(2.2-2.7)	1.96(1.29-2.96)	1.80(1.20-2.69)
East Midlands	358	100,056	3.6(3.2-4.0)	2.90(1.92-4.38)	3.04(2.04-4.54)
West Midlands	747	324,145	2.3(2.1-2.5)	1.87(1.25-2.80)	2.01(1.35-2.97)
East of England	818	351,442	2.3(2.2-2.5)	1.89(1.26-2.83)	2.27 (1.53-3.36)
South West	667	313,216	2.1(2.0-2.3)	1.73(1.15-2.59)	1.72(1.16-2.55)
South Central	566	321,555	1.8(1.6-1.9)	1.43(0.95-2.15)	1.77(1.19-2.63)
London	620	364,800	1.7(1.6-1.8)	1.38(0.92-2.07)	1.87(1.26-2.78)
South East Coast	770	300,909	2.6(2.4-2.7)	2.08(1.39-3.11)	2.75(1.86-4.08)

Table 39 Univariable and Multivariable multilevel Poisson analyses of GP sore throat consultation in adults 16-44 years old from CALIBER database

Analysis of sore throat in adults							
Characteristic	Rate date			Univariable analysis		Multivariable analysis	
	Sore throat consultations	Person years	Incidence/1000 person years (95%CI)	Incidence Rate Ratio(95%CI)	P	Adjusted IRR	P
Gender							
Male	238,255	3,897,287	61(61-61)	1	<0.001	1	<0.001
Female	462,300	3,835,896	121(120-121)	1.97(1.96-1.99)		1.75(1.71-1.79)	
Age category							

16-24 years	268,000	2,316,067	116(115-116)	1	<0.001	1	<0.001
25-44 years	432,555	5,417,116	80(80-80)	0.69(0.69-0.70)		0.51(0.49-0.52)	
Ethnic origin							
White British	321,006	2,662,499	121(120-122)	1	<0.001	1	<0.001
Indian	4,681	41,891	112(107-116)	0.93(0.89-0.97)		1.06(0.98-1.14)	
Black African	3,020	33,515	90(86-95)	0.75(0.71-0.79)		0.83(0.72-0.95)	
Black Caribbean	2,617	26,503	99(93-105)	0.82(0.77-0.87)		0.87(0.79-0.97)	
Black other	1,771	19,983	89(83-95)	0.74(0.69-0.79)		0.85(0.75-0.95)	
Bangladeshi	1,028	8,297	124(113-136)	1.03(0.94-1.13)		1.12(0.97-1.29)	
Pakistani	4,012	28,491	141(134-148)	1.17(1.11-1.23)		1.22(1.13-1.32)	
Other Asian	1,752	17,802	98(92-105)	0.82(0.76-0.88)		0.97(0.86-1.10)	
Chinese	772	9,982	77(68-86)	0.64(0.58-0.71)		0.78(0.68-0.90)	
Mixed	2,146	21,632	99(93-106)	0.82(0.77-0.88)		0.80(0.71-0.89)	
Other	5,718	55,433	103(99-107)	0.86(0.82-0.89)		0.99(0.93-1.05)	
Unknown	100,938	937,737	107(107-109)	0.89(0.88-0.90)		0.96(0.93-0.98)	
Social Deprivation							
Least deprived	108,813	1,225,727	89(88-90)	1	<0.001	1	0.01

2nd least deprived	138,642	1,565,682	89(88-90)	1.00(0.99-1.01)		1.01(0.95-1.07)	
3rd least deprived	133,300	1,455,246	92(91-92)	1.03(1.02-1.04)		1.02(0.95-1.10)	
4th least deprived	154,525	1,697,991	91(90-92)	1.03(1.01-1.04)		1.02(0.95-1.10)	
Most deprived	161,441	1,745,740	92(92-93)	1.04(1.03-1.05)		1.09(1.00-1.19)	
Respiratory illness							
Absent	525,725	6,268,955	84(84-84)	1	<0.001	1	<0.001
Present	174,830	1,454,228	119(119-120)	1.42(1.41-1.44)		1.26(1.23-1.29)	
Obstructive sleep apnoea							
Absent	698,256	7,717,586	90(90-91)	1	<0.001	1	<0.001
Present	2,299	15,597	147(139-157)	1.63(1.53-1.73)		1.45(1.31-1.61)	
Obesity							
Not coded	626,511	7,266,361	86(86-87)	1	<0.001	1	<0.001
Obese	74,044	466,822	159(157-160)	1.84 (1.82-1.86)		1.45(1.26-1.43)	
HIV status							
HIV negative	697,306	7,707,937	90(90-91)	1	<0.001	1	<0.001
HIV positive	3,249	25,246	129(122-136)	1.43(1.35-1.50)		1.37(1.24-1.51)	

Eating disorder							
Absent	691,868	7,668,696	90(89-91)	1	<0.001	-	-
Eating disorder	8,687	64,488	135(131-139)	1.49(1.45-1.54)		-	
Alcohol							
Non-drinker	101,557	920,847	110(109-111)	1	<0.001	1	<0.001
Mild-Moderate drinker	160,772	1,411,135	114(113-115)	1.03(1.02-1.05)		1.03(0.99-1.07)	
Heavy drinker	9,052	166,358	54(53-56)	0.49(0.48-0.51)		0.65(0.60-0.71)	
Smoking							
Non-Smoker	384,311	4,941,860	78(77-78)	1	<0.001	1	<0.001
Ex-smoker	59,408	510,992	116(115-118)	1.49(1.48-1.51)		1.17(1.14-1.21)	
Smoker	256,836	2,280,331	113(112-113)	1.45(1.44-1.46)		1.10(1.08-1.12)	
Diabetes							
No Diabetes coded	690,038	7,661,056	90(88-90)	1	<0.001	1	<0.001
Diabetes coded	10,517	72,127	146(142-150)	1.62(1.57-1.67)		1.20(1.15-1.26)	
Hypertension							

No hypertension coded	668,963	7,524,266	89(89-89)	1	<0.001	1	<0.001
Hypertension coded	31,592	208,917	151(149-154)	1.70(1.67-1.73)		1.44(1.39-1.50)	
Practice region							
North East	4,751	58,266	82(78-85)	1	<0.001	1	0.01
North West	118,007	1,208,414	98(97-99)	1.20(1.15-1.25)		1.13(0.84-1.52)	
Yorkshire & The Humber	32,244	354,395	91(89-93)	1.12(1.07-1.17)		1.07(0.77-1.49)	
East Midlands	30,143	297,100	101(100-103)	1.24(1.19-1.30)		1.18(0.86-1.61)	
West Midlands	101,554	966,445	105(104-106)	1.29(1.24-1.34)		1.23(0.92-1.64)	
East of England	98,107	1,043,158	94(93-95)	1.15(1.11-1.20)		1.14(0.83-1.54)	
South West	85,165	893,425	95(94-96)	1.17(1.12-1.22)		1.12(0.82-1.52)	
South Central	75,475	903,698	84(83-84)	1.02(0.98-1.07)		0.98(0.72-1.32)	
London	84,002	1,176,780	71(71-72)	0.88(0.84-0.91)		0.97(0.71-1.31)	
South East Coast	71,107	831,503	86(85-86)	1.05(1.01-1.09)		0.99(0.73-1.36)	

Table 40 Univariable and Multivariable multilevel Poisson analyses of recurring sore throat (measured by 3 GP GP sore throat consultation in 12 months) in adults 16-44 years old from CALIBER database

Analysis of rec sore throat in adults							
Characteristic	Rate date			Univariable analysis		Multivariable analysis	
	Patients with rec sore throat	Person years	Incidence/1000 person years (95%CI)	Incidence Rate Ratio(95%CI)	P	Adjusted IRR	P
Gender							
Male	5,506	3,872,167	1.4(14-15)	1	<0.001	1	<0.001
Female	14,917	3,773,283	4.0(39-40)	2.78(2.70-2.87)		1.98(1.82-2.15)	
Age category							
16-24 years	5,939	2,065,130	2.9(28-29)	1	<0.001	1	<0.001
25-44 years	14,484	5,580,320	2.6(26-26)	0.90(0.88-0.93)		0.66(0.61-0.71)	
Ethnic origin							
White British	10,571	2,615,456	4.0(40-41)	1	<0.001	1	0.004
Indian	144	41,331	3.5(30-41)	0.86(0.73-1.02)		0.95(0.74-1.23)	
Black African	81	33,289	2.4(20-30)	0.60(0.48-0.75)		0.68(0.44-1.03)	

Black Caribbean	62	26,239	2.4(18-30)	0.58(0.48-0.75)		0.61(0.45-0.83)	
Black other	50	19,807	2.5(19-33)	0.62(0.47-0.82)		0.78(0.58-1.03)	
Bangladeshi	37	8,179	4.5(33-62)	1.12(0.81-1.54)		1.14(0.75-1.73)	
Pakistani	141	27,984	5.0(43-59)	1.25(1.06-1.47)		1.18(0.99-1.40)	
Other Asian	63	17,626	3.6(30-46)	0.88(0.69-1.13)		1.04(0.74-1.45)	
Chinese	23	9,912	2.3(15-35)	0.57(0.38-0.86)		0.69(0.43-1.13)	
Mixed	72	21,367	3.4(27-42)	0.83(0.66-1.05)		0.92(0.69-1.24)	
Other	173	54,721	3.2(27-37)	0.78(0.67-0.91)		0.82(0.65-1.05)	
Unknown	3,063	924,495	3.3(32-34)	0.82(0.79-0.85)		0.89(0.82-0.96)	
Social Deprivation							
Least deprived	3,263	1,211,570	2.6(26-28)	1	0.69	-	
2nd least deprived	4,084	1,548,093	2.6(26-27)	0.98(0.94-1.03)		-	
3rd least deprived	3,906	1,438,275	2.7(26-28)	1.01(0.96-1.06)		-	
4th least deprived	4,446	1,678,797	2.6(26-27)	0.98(0.94-1.03)		-	
Most deprived	4,616	1,726,382	2.7(26-28)	0.99(0.95-1.04)		-	
Respiratory illness							
Absent	14,431	6,208,400	2.3(23-24)	1	<0.001	1	<0.001

Present	5,992	1,437,050	4.2(41-43)	1.79(1.74-1.85)		1.53(1.44-1.63)	
Obstructive sleep apnoea							
Absent	20,350	7,630,143	2.7(26-27)	1	<0.001	1	<0.001
Present	73	15,308	4.8(38-60)	1.79(1.42-2.25)		1.82(1.34-2.48)	
Obesity							
Not coded	17,880	7,190,495	2.5(25-25)	1	<0.001	1	<0.001
Obese	2,543	454,956	5.6(54-58)	2.25(2.16-2.34)		1.55(1.40-1.71)	
HIV status							
HIV negative	20,323	7,620,653	2.7(26-27)	1	<0.001	-	
HIV positive	100	24,798	4.0(33-49)	1.51(1.24-1.84)		-	
Eating disorder							
Absent	20,123	7,582,251	2.7(26-27)	1	<0.001	-	
Eating disorder	300	63,199	4.7(42-53)	1.79(1.60-2.00)		-	
Alcohol							
Non-drinker	3,236	907,956	3.6(34-37)	1	<0.001	1	<0.001

Mild-Moderate drinker	4,881	1,388,452	3.5(34-36)	0.99(0.94-1.03)		0.94(0.87-1.01)	
Heavy drinker	299	165,285	1.4(12-16)	0.39(0.33-0.44)		0.48(0.40-0.58)	
Smoking							
Non-Smoker	10,877	4,896,800	2.2(22-23)	1	<0.001	1	0.003
Ex-smoker	1,599	504,568	3.2(30-33)	1.43(1.35-1.50)		1.06(0.96-1.17)	
Smoker	7,947	2,244,082	3.5(35-36)	1.59(1.55-1.64)		1.12(1.06-1.18)	
Diabetes							
No Diabetes coded	20,115	7,574,558	2.7(26-27)	1	<0.001	-	
Diabetes coded	308	70,892	4.3(39-49)	1.64(1.46-1.83)		-	
Hypertension							
No hypertension coded	19,534	7,440,058	2.6(26-27)	1	<0.001	1	<0.001
Hypertension coded	889	205,392	4.3(41-46)	1.65(1.54-1.76)		1.26(1.13-1.41)	
Practice region							
North East	128	57,664	2.2(19-26)	1	<0.001	1	0.002
North West	3,541	1,191,705	3.0(29-31)	1.34(1.12-1.60)		1.28(0.87-1.90)	

Yorkshire & The Humber	956	350,530	2.7(26-29)	1.23(1.02-1.48)	1.23(0.79-1.92)
East Midlands	978	292,925	3.3(31-36)	1.50(1.25-1.81)	1.54(1.02-2.31)
West Midlands	3,130	952,212	3.3(32-34)	1.48(1.24-1.77)	1.44(0.98-2.10)
East of England	2,948	1,030,596	2.9(28-30)	1.29(1.08-1.54)	1.19(0.80-1.78)
South West	2,507	882,597	2.8(27-30)	1.28(1.07-1.54)	1.18(0.79-1.78)
South Central	2,019	895,430	2.3(22-24)	1.02(0.85-1.21)	0.89(0.60-1.32)
London	2,246	1,168,208	1.9(18-20)	0.87(0.72-1.04)	1.01(0.67-1.51)
South East Coast	1,970	823,582	2.4(23-25)	1.08(0.90-1.29)	1.03(0.69-1.54)

Analysis of tonsillectomy in adults							
Characteristic	Rate date			Univariable analysis		Multivariable analysis	
	Tonsillectomies	Person years	Incidence/1000 person years (95%CI)	Incidence Rate Ratio(95%CI)	P	Adjusted IRR	P
Gender							
Male	2,162	3,886,193	0.6(5-6)	1	<0.001	1	<0.001

Female	4,668	3,814,292	1.2(12-13)	2.20(2.09-2.31)		1.39(1.24-1.55)	
Age category							
16-24 years	3,771	2,079,749	1.8(18-19)	1	<0.001	1	<0.001
25-44 years	3,059	5,620,736	0.5(5-6)	0.30(0.29-0.31)		0.16(0.14-0.17)	
Ethnic origin							
White British	5,115	2,640,103	1.9(19-20)	1	<0.001	1	<0.001
Indian	43	41,677	1.0(8-14)	0.53(0.39-0.72)		0.85(0.54-1.34)	
Black African	22	33,444	0.7(4-10)	0.34(0.22-0.52)		0.40(0.23-0.69)	
Black Caribbean	30	26,378	1.1(8-16)	0.59(0.41-0.84)		0.66(0.37-1.18)	
Black other	20	19,922	1.0(6-16)	0.52(0.33-0.80)		0.53(0.28-0.99)	
Bangladeshi	16	8,257	1.9(12-32)	1.00(0.61-1.63)		1.63(1.01-2.64)	
Pakistani	35	28,355	1.2(9-17)	0.64(0.46-0.89)		0.85(0.52-1.40)	
Other Asian	23	17,735	1.3(9-20)	0.67(0.44-1.00)		1.16(0.50-2.72)	
Chinese	5	9,965	0.5(2-12)	0.26(0.11-0.62)		0.57(0.21-1.52)	
Mixed	42	21,487	2.0(14-26)	1.01(0.74-1.37)		0.79(0.50-1.27)	
Other	64	55,178	1.2(9-15)	0.60(0.47-0.77)		0.57(0.39-0.85)	
Unknown	1,415	931,080	1.5(14-16)	0.78(0.74-0.83)		0.79(0.70-0.89)	

Social Deprivation							
Least deprived	1,125	1,220,412	0.9(9-10)	1	<0.001	1	0.67
2nd least deprived	1,553	1,558,538	1.0(9-10)	1.08(1.00-1.17)		1.07(0.93-1.21)	
3rd least deprived	1,349	1,448,632	0.9(9-10)	1.01(0.93-1.09)		1.05(0.91-1.20)	
4th least deprived	1,427	1,690,967	0.8(8-9)	0.92(0.85-0.99)		1.11(0.97-1.27)	
Most deprived	1,347	1,739,316	0.7(7-8)	0.84(0.77-0.91)		1.06(0.92-1.22)	
Respiratory illness							
Absent	4,955	6,245,702	0.8(8-8)	1	<0.001	1	<0.001
Present	1,875	1,454,784	1.3(12-13)	1.62(1.54-1.71)		1.18(1.07-1.29)	
Obstructive sleep apnoea							
Absent	6,776	7,685,079	0.9(9-9)	1	<0.001	1	<0.001
Present	54	15,406	3.5(27-46)	3.98(3.04-5.20)		2.59(1.59-4.23)	
Obesity							
Not coded	6,083	7,237,423	0.8(8-9)	1	<0.001	1	<0.001
Obese	747	463,063	1.6(15-17)	1.92(1.78-2.07)		1.58(1.39-1.80)	
HIV status							

HIV negative	6,805	7,675,332	0.9(9-9)	1	0.57	-	
HIV positive	25	25,153	1.0(7-15)	1.12(0.76-1.66)		-	
Eating disorder							
Absent	6,732	7,636,454	0.9(9-9)	1	<0.001	-	
Eating disorder	98	64,031	1.5(13-19)	1.74(1.42-2.12)		-	
Alcohol							
Non-drinker	880	916,718	1.0(9-10)	1	<0.001	1	<0.001
Mild-Moderate drinker	1,513	1,404,039	1.1(10-11)	1.12(1.03-1.22)		1.26(1.14-1.40)	
Heavy drinker	71	166,006	0.4(3-5)	0.45(0.35-0.57)		0.61(0.47-0.78)	
Smoking							
Non-Smoker	3,385	4,925,188	0.7(6-7)	1	<0.001	1	<0.001
Ex-smoker	527	508,724	1.0(10-11)	1.51(1.38-1.65)		1.45(1.24-1.69)	
Smoker	2,918	2,266,573	1.3(12-13)	1.87(1.78-1.97)		1.30(1.20-1.42)	
Diabetes							
No Diabetes coded	6,755	7,628,751	0.9(9-9)	1	0.15	-	
Diabetes coded	75	71,734	1.0(8-13)	1.18(0.94-1.48)		-	

Hypertension							
No hypertension coded	6,618	7,492,583	0.9(9-9)	1	0.04	-	
Hypertension coded	212	207,902	1.0(9-12)	1.15(1.01-1.32)		-	
Practice region							
North East	63	57,950	1.1(8-14)	1	<0.001	1	0.38
North West	1317	1,201,895	1.1(10-12)	1.01(0.78-1.30)		0.88(0.57-1.38)	
Yorkshire & The Humber	382	352,544	1.1(10-12)	1.00(0.76-1.30)		0.97(0.59-1.59)	
East Midlands	338	295,318	1.3(12-15)	1.21(0.93-1.58)		1.23(0.75-2.02)	
West Midlands	886	961,959	0.9(9-10)	0.85(0.66-1.09)		0.91(0.59-1.40)	
East of England	904	1,038,785	0.9(8-9)	0.80(0.62-1.03)		0.92(0.59-1.45)	
South West	910	888,995	1.0(10-11)	0.94(0.73-1.22)		0.89(0.58-1.39)	
South Central	673	900,495	0.7(7-8)	0.69(0.53-0.89)		0.90(0.56-1.42)	
London	597	1,174,248	0.5(5-6)	0.47(0.36-0.61)		0.78(0.50-1.22)	
South East Coast	710	282,297	0.9(8-9)	0.79(0.61-1.02)		0.86(0.54-1.35)	

Comparison of risk factors across settings, from sore throat in primary care through to tonsillectomy in secondary care						
Characteristic	Children			Adults		
	Sore throat (adjusted IRR)	Rec sore throat (adjusted IRR)	Tonsillectomy (adjusted IRR)	Sore throat (adjusted IRR)	Rec sore throat (adjusted IRR)	Tonsillectomy (adjusted IRR)
Gender						
Male	1	1	1	1	1	1
Female	1.24(1.23-1.26)	1.40(1.36-1.45)	1.10(1.05-1.16)	1.75(1.71-1.79)	1.98(1.82-2.15)	1.39(1.24-1.55)
Age category						
0-4 years	1	1	1	-	-	-
5-15 years	0.55(0.55-0.56)	0.64(0.62-0.67)	5.34(4.64-6.15)	-	-	-
16-24 years	-	-	-	1	1	1
25-44 years	-	-	-	0.51(0.49-0.52)	0.66(0.61-0.71)	0.16(0.14-0.17)
Ethnic origin						
White British	1	1	1	1	1	1
Indian	0.87(0.82-0.93)	0.67(0.55- 0.82)	0.86(0.66-1.13)	1.06(0.98-1.14)	0.95(0.74-1.23)	0.85(0.54-1.34)
Black African	0.63(0.59-0.68)	0.42(0.33- 0.54)	0.71(0.52-0.97)	0.83(0.72-0.95)	0.68(0.44-1.03)	0.40(0.23-0.69)
Black Caribbean	0.68(0.62-0.74)	0.50(0.38- 0.67)	0.50(0.34-0.74)	0.87(0.79-0.97)	0.61(0.45-0.83)	0.66(0.37-1.18)

Black other	0.69(0.63-0.75)	0.61(0.47- 0.81)	0.86(0.62-1.20)	0.85(0.75-0.95)	0.78(0.58-1.03)	0.53(0.28-0.99)
Bangladeshi	1.01(0.90-1.13)	1.12(0.84-1.48)	0.61(0.35-1.06)	1.12(0.97-1.29)	1.14(0.75-1.73)	1.63(1.01-2.64)
Pakistani	0.96(0.91-1.02)	0.90(0.76- 1.06)	1.01(0.80-1.28)	1.22(1.13-1.32)	1.18(0.99-1.40)	0.85(0.52-1.40)
Other Asian	0.82(0.75-0.89)	0.70(0.54- 0.90)	0.94(0.65-1.34)	0.97(0.86-1.10)	1.04(0.74-1.45)	1.16(0.50-2.72)
Chinese	0.72(0.62-0.82)	0.38(0.21- 0.68)	0.47(0.21-1.04)	0.78(0.68-0.90)	0.69(0.43-1.13)	0.57(0.21-1.52)
Mixed	0.67(0.63-0.71)	0.45(0.37- 0.55)	0.92(0.73-1.16)	0.80(0.71-0.89)	0.92(0.69-1.24)	0.79(0.50-1.27)
Other	0.85(0.81-0.89)	0.79(0.68- 0.92)	0.89(0.73-1.08)	0.99(0.93-1.05)	0.82(0.65-1.05)	0.57(0.39-0.85)
Unknown	0.92(0.90-0.93)	0.83(0.80- 0.87)	0.62(0.58-0.66)	0.96(0.93-0.98)	0.89(0.82-0.96)	0.79(0.70-0.89)
Social Deprivation						
Least deprived	1	1	1	1	-	1
2nd least deprived	1.02(1.00-1.04)	1.03(0.97-1.08)	0.98(0.90-1.06)	1.01(0.95-1.07)	-	1.07(0.93-1.21)
3rd least deprived	1.02(1.00-1.04)	1.06(0.01-1.12)	0.88(0.81-0.96)	1.02(0.95-1.10)	-	1.05(0.91-1.20)
4th least deprived	0.97(0.95-0.97)	0.93(0.88-0.98)	0.76(0.70-0.82)	1.02(0.95-1.10)	-	1.11(0.97-1.27)
Most deprived	0.94(0.92-0.96)	0.89(0.84-.094)	0.68(0.62-0.74)	1.09(1.00-1.19)	-	1.06(0.92-1.22)
Respiratory illness						
Absent	1	1	1	1	1	1
Present	1.79(1.76-1.81)	2.03(1.95-2.10)	1.50(1.42-1.59)	1.26(1.23-1.29)	1.53(1.44-1.63)	1.18(1.07-1.29)

Obstructive sleep apnoea						
Absent	1	1	1	1	1	1
Present	1.79(1.76-1.81)	2.10(1.69-2.63)	10.99(9.7-12.40)	1.45(1.31-1.61)	1.82(1.34-2.48)	2.59(1.59-4.23)
Obesity						
Not coded	1	1	1	1	1	1
Obese	2.00(1.93-2.08)	2.31(2.12-2.53)	1.96(1.66-2.30)	1.45(1.26-1.43)	1.55(1.40-1.71)	1.58(1.39-1.80)
HIV status						
HIV negative	1	-	-	1	-	-
HIV positive	1.30(1.17-1.45)	-	-	1.37(1.24-1.51)	-	-
Eating disorder						
Absent	1	1	1	-	-	-
Eating disorder	1.88(1.76-2.01)	2.30(1.95-2.70)	1.54(1.13-2.12)	-	-	-
Alcohol	-	-	-			
Non-drinker	-	-	-	1	1	1
Mild-Moderate drinker	-	-	-	1.03(0.99-1.07)	0.94(0.87-1.01)	1.26(1.14-1.40)

Heavy drinker	-	-	-	0.65(0.60-0.71)	0.48(0.40-0.58)	0.61(0.47-0.78)
Smoking	-	-	-			
Non-Smoker	-	-	-	1	1	1
Ex-smoker	-	-	-	1.17(1.14-1.21)	1.06(0.96-1.17)	1.45(1.24-1.69)
Smoker	-	-	-	1.10(1.08-1.12)	1.12(1.06-1.18)	1.30(1.20-1.42)
Diabetes	-	-	-			
No Diabetes coded	-	-	-	1	-	-
Diabetes coded	-	-	-	1.20(1.15-1.26)	-	-
Hypertension	-	-	-			
No hypertension coded	-	-	-	1	1	-
Hypertension coded	-			1.44(1.39-1.50)	1.26(1.13-1.41)	-

Appendix E Search strategy for tonsillectomy outcomes

Pubmed

1. “Tonsillectomy” [Mesh] OR tonsillectom* [ti] OR tonsilectom* [ti] OR adenotonsillectom* [ti] OR adeno-tonsillectom* [ti]
2. "Palatine Tonsil/surgery"[Mesh]
3. (Tonsil* [ti] OR adenotonsil* [ti]) AND (SURG* [ti] OR OPERAT* [ti] OR EXCIS* [ti] OR EXTRACT* [ti] OR REMOV* [ti] OR DISSECT* [ti] OR ABLAT* [ti] OR COBLAT* [ti] OR LASER* [ti])
4. #1 OR #2 OR #3

EMBASE (Ovid)

1. exp *tonsillectomy/
2. exp tonsil/su [Surgery]
3. (tonsillectom* or tonsilectom* or adenotonsillectom* or adeno-tonsillectom*).ti.
4..((Tonsil* or adenotonsil*) and (SURG* or OPERAT* or EXCIS* or EXTRACT* or REMOV* or DISSECT* or ABLAT* or COBLAT* or LASER*)).ti.
5. 1 or 2 or 3 or 4

CINAHL (EBSCO)

S1 (MH "Tonsillectomy")

S2 TI tonsillectom* OR tonsilectom* OR adenotonsillectom* OR adeno-tonsillectom*

S3 TI (tonsil* OR adenotonsil*) AND (surg* OR laser* OR extract* OR resect* OR excis* OR operat* OR dissect* OR remov* OR coblat* OR ablat*)

S4 S1 or S2 or S3

Appendix F – Critical appraisal of tonsillectomy outcome studies using CASP and Summary

Outcome: Days of sore throat

Type of study: Meta-analysis Year of publication: 2014

What did the study involve? Meta-analysis of 2 RCTs investigating effectiveness of tonsillectomy in adults with recurring tonsillitis.

What was the risk of bias? Inappropriate selection of patients (heterogenous), insufficient follow up (only 6 months)

What were the results? Pooled mean difference for number of days with sore throat in a follow-up period of about six months was 10.6 days fewer in favour of the group receiving surgery (95% CI 5.8 fewer to 15.8 fewer).

Are the results relevant to my study? The definition used for recurring sore throat (3 episodes of pharyngitis in 6 months or 3 episodes of streptococcal pharyngitis in 12 months) seems less stringent to the one I used (7 cases of self-reported tonsillitis in 12 months, or 5 episodes/year over 2 years or 3 episodes/year over 3 years).

Outcome: Episodes of sore throat

Type of study: Meta-analysis

Year of publication: 2014

What did the study involve? Meta-analysis of 2 RCTs investigating effectiveness of tonsillectomy in adults with recurring tonsillitis.

What was the risk of bias? Inappropriate selection of patients (heterogenous), insufficient follow up (only 6 months)

What were the results? The pooled results of the two adult studies (n=156) showed there were 3.6 fewer sore throat episodes in the group receiving tonsillectomy in the first 6 months after treatment (95% CI 7.9 fewer to 0.70 more).

Are the results relevant to my study? The definition used for recurring sore throat (3 episodes of pharyngitis in 6 months or 3 episodes of streptococcal pharyngitis in 12 months) seems less stringent to the one I used (7 cases of self-reported tonsillitis in 12 months, or 5 episodes/year over 2 years or 3 episodes/year over 3 years).

Outcome: Quality of life

Type of study: Systematic review

Year of publication: 2013

What did the study involve? Review eight studies investigating the role of tonsillectomy on the quality of life for adults with recurring tonsillitis.

What was the risk of bias? The review authors used only two search engines to identify studies and as a result missed two further studies identified through my search strategy (Powell, Skevas). The reported results are based on studies with very low response rates (4 studies had less than 50% response rate) and the implication on selection bias. Considerable heterogeneity between studies

What were the results? Six used the Glasgow Benefit inventory (GBI) and two used Short Form questionnaires (SF12 and SF36). The GBI scale of –100 to +100, positive values indicate quality of life (QoL) improvement, measures perceived improvement following a treatment and is

administered only once. Six studies that used this tool had between 47 and 187 participants with response rates of 30-89%. Total GBI scores ranged from 15.78 to 35.2. Finally, the ages of participants varied considerably from 15-25 or 15-60. One study used the SF36 at one year after tonsillectomy whilst another used the SF12 at 6 and 12 months after tonsillectomy. Response rates were between 97% and 56%. Studies reported an improvement in the physical component of quality of life (7.6-10.1).

Are the results relevant to my study? Definitions of recurring/chronic tonsillitis (e.g. 3 episodes in 12 months) were not always similar to those used in my study (e.g. 7episodes in 12 months).

Outcome: GP visits for sore throat

Type of study: RCT

Year of publication: 2012

What did the study involve? Randomised controlled trial of 86 adults with recurring pharyngitis and reported on GP visits as a secondary outcome.

What was the risk of bias? Authors used GP visits as a secondary outcome. This was measured through participant self-reports on days off work, upto 5 months after randomisation. The authors do not comment as to whether they were powered to study this.

What were the results? Four percent pf participants who had a tonsillectomy visited their GP for pharyngitis whereas 43% of the control group did so in the five months after randomisation (difference 38%, 95% CI 22% to 55%).

Are the results relevant to my study? This study was conducted on a Finnish population of patients 13 years and older with recurring pharyngitis (>2 episodes in 12 months) and care must be taken when translating the findings to our population of interest (that is adults over 15 from England who had 7 episodes of tonsillitis over the preceding year). Different healthcare systems could affect whether illness results in GP visits and care must be taken in interpreting these results for our study population.

Outcome: Snoring reduction

Type of study: Retrospective cohort study

Year of publication: 2008

What did the study involve? Cohort comprised of 460 adults recruited from ENT clinics (family members of patients) and hospital employee lists. Respondents were asked to recall if they had had a tonsillectomy and whether they currently snore. The authors chose to define their cases – habitual snorers – as those who reported their frequency of snoring as always or every night. The graded habitual snoring severity as mild, moderate (rarely irritates other people), or severe (roommates choose to sleep in another room). They graded the risk of being a habitual snorer of participants who'd had a tonsillectomy compared to those who hadn't.

What was the risk of bias? Description of the cohort recruitment is not supplied and therefore it is difficult to conclude on the suitability of their recruitment method. Especially, since more than half of their cohort had received a tonsillectomy, which is much higher than in the general population. This data was recorded from patient subjective reports and may not be as accurate as data collected from patient health records, which may have provided more information as to when the procedure happened and its original indicated purpose.

What were the results? After the authors controlled for confounding factors such as age, sex, and body mass index, they found that not receiving a tonsillectomy increased the risk of being a habitual snorer (OR 1.18, 95% CI 1.15-2.86).

Are the results relevant to my study? This study using Turkish adults (18 and over), sourced from ENT clinics and hospitals, who had a 50% prevalence rate of tonsillectomy in their population, do not seem comparable to the UK general public, where the rate is less than half that.

Outcome: Change in voice

Type of study: Case-control

Year of publication: 2009

What did the study involve? Cases were recruited from the local hospital, whilst controls were recruited from a nearby school at a ratio of 1:2. Controls were age and sex matched to cases. The authors used objective measures of voice: fundamental frequency, jitter, shimmer, harmonic, noise ratio, long-term average spectrum, and nasalance, which were assessed preoperatively and 4 weeks after tonsillectomy in the case-group and once only in the control group.

What was the risk of bias? There was no mention of criteria used to select cases, or what the recruitment rate was. Additionally, the age structure of the respondents is not reported. Loss to follow up numbers aren't reported.

What were the results? The authors reported that whilst hypernasality reduced after tonsillectomy, compared to preoperative readings, there was no statistical difference in any component of voice measured between pre-operative and post-operative measurements, or between those that received tonsillectomy and those that did not.

Are the results relevant to my study? This is an Indian population with no report of how cases were selected so applicability to our dataset is difficult to interpret.

Outcome: Immunological profile

Type of study: Case series

Year of publication: 1996

What did the study involve? The population randomly recruited included adults listed for a tonsillectomy to treat recurring or chronic tonsillitis in the local Finnish ENT department. Authors measured pre-operative saliva to quantify markers of immunity. Post-operatively the authors took saliva samples at 1 and 6 months after tonsillectomy.

What was the risk of bias? Selection bias as the patients were not sequentially recruited. Detection bias as measurement used may miss significant change in immune system – that is increased rate of infections.

What were the results? Authors analysed saliva for selected host defence factors, representing both immune (total IgA, IgG, IgM, anti-Streptococcus mutans, anti-EBV, anti-CMV, and anti-adenovirus IgA and IgG) and nonimmunoglobulin (lysozyme, lactoferrin, salivary peroxidases, thiocyanate, hypothiocyanite, and agglutinins) mediators. Following tonsillectomy, a significant ($P < 0.04$) reduction was observed in specific IgG antibodies, suggesting that tonsils participate in local IgG response to oral antigens. Total IgM levels also decreased ($P< 0.006$), which may to some extent reflect reduced antigenic stimuli compared to preoperative status with frequent tonsillitis. Saliva-derived

nonimmunoglobulin host defence factors, except lactoferrin, which declined significantly, remained normal throughout the study period. The authors concluded that tonsillectomy does not seem to lead to any significant long-term impairment of salivary defence capacity.

Are the results relevant to my study? There is insufficient evidence on the study sample to make reasonable judgments about the generalisability of their results to our study population. Additionally, the lack of clinical correlation with their findings makes the translation of this information for our study population almost meaningless.

Outcome: Haemorrhage risk

Type of study: Case series

Year of publication: 2005

What did the study involve? The Royal College of Surgeons of England undertook an audit of all tonsillectomies undertaken in England and Northern Ireland between 2002-2004 to ascertain complication rates. This national audit that captured nearly every tonsillectomy undertaken in England between 2002-2004 and included 76% of all tonsillectomies undertaken that during that period (i.e. 33,921 patients).

What was the risk of bias? Detection bias as it is unable to define how many patients managed post-tonsillectomy bleeding without hospital presentation.

What were the results? The report suggested that there was an overall post-operative haemorrhage rate of 3.5%.

Are the results relevant to our study? Yes since this is a large UK population sample of those underwent tonsillectomy.

Outcome: Halitosis

Type of study: Case series

Year of publication:

What did the study involve? The study recruited 44 patients with halitosis and chronic tonsillitis. The authors measured halitosis at 4 and 8 weeks post-operatively using Finklesteins test.

What was the risk of bias? Selection bias as no report of how patients were selected. Detection bias as test surgeon who performed operation is subjectively reporting whether tonsillectomy improved halitosis.

What were the results? He used subjective measures of halitosis at 4 and 8 weeks following tonsillectomy and reported 79.5% of patients reported improved symptoms.

Are the results relevant to our study? Nigerian population with no description of cohort or inclusion criteria to difficult to ascertain relevance to our population.

Outcome: Taste

Type of study: Case series

Year of publication: 2010

What did the study involve? The authors recruited 60 adults who did not have a history of olfactory or gustatory disorder. The authors tested patients' ability to discriminate between four tastants on four different regions of the tongue on the 1st post-operative day, 15 days and then again at one month.

What was the risk of bias? Selection bias as patients were not reported as consecutive. Detection bias as insufficient follow up to define true incidence of complication

What were the results? The final evaluation, 1 month postoperatively, yielded normal results for all the patients except one. Presentation of low stimulus quantities resulted in recognition percentages of 95%, while higher stimulus quantities elicited correct responses by all of the patients, except one. Results were not reported with measures of precision (standard error, confidence intervals), which brings their reliability into question.

Are the results relevant to our study? Greek population therefore may be different from our study population.

Outcome: Societal cost

Type of study: Economic

Year of publication: 2002

What did the study involve? This study used a postal survey methodology to assess economic burden. This study used the Glasgow Benefit Inventory (GBI) to measure quality of life and thus calculate quality adjusted life years. The study used the costs of surgery, antibiotics, work days missed and physician visits of tonsillitis to calculate a break-even point. They failed to incorporate costs of far more commonly used analgesics

What was the risk of bias? The authors report an extremely low response rate (response rate<30%), and care should be taken in interpreting results due to selection bias

Are the results relevant to our study? The authors reported that for adults with recurring tonsillitis, tonsillectomy has a high up-front cost, but within 2.3 years there is no difference in costs between having and not having the operation.

Are the results relevant to our study? This study conducted on adults with recurring tonsillitis in USA depends, calculates its results based on local healthcare costs. There is considerable difference in healthcare costs between England and USA, and the results are not easily generalisable to our study population.

List of factors	Type of study (level of evidence)	Specific (non-standardised) effect size	Standardised Effect small/med/large	Potential impact on decision making
Days of sore throat(54)	Meta-analysis of RCT (1a)	3.61/6months	1. large	1. Strong
Number of episodes of sore throat(54)	Meta-analysis of RCT (1a)	10.64/6months	1. large	1. Strong
Visits to the GP(7)	RCT (1b)	0.9/6months	1. large	1. Strong
QoL(385)	Systematic review of cohort (2a)	10% improvement in physical component sf 36	2. medium	1. Moderate
Snoring reduction(386)	Case series (4)	46% reduction in odds of being a severe/habitual snorer if you had a tonsillectomy	1. large	2. Weak
Voice change (387)	Case control (3b)	17% of peripubescent males had hypernasalance preop, and 9% had it post tonsillectomy	2. small	3. Weak
Halitosis reduction(388)	Case series (4)	80% had clearance of halitosis at 2 months	1. large	3. Weak
Societal cost(8)	Case series (4)	12.3 year break even time	2. medium	3. Weak
Taste disturbance(389)	Case series (4)	10%/6months	2. medium	3. Weak
Haemorrhage risk(56)	Cohort (2b)	4.9% incidence adults (all indication tonsillectomy)	1. large	1.Strong

Immunological profile(390)	Case series (4)	Mild reduction in IgM	3. small	3. Weak

Table 41 Critical appraisal of knowledge related to eleven treatment outcomes for recurring sore throat

Appendix G – Online ranking exercise

Exercise

Please rank, in order of importance, which factors about a treatment and outcome are most important to you when deciding how best to treat recurring sore throat? (Order the list from 1-10, with 1 meaning most important for you in choosing a treatment, and 10 being the least important to you)

Factor	Your rank (1-10) of importance in your decision
Reducing days of sore throat	
Reducing unexpected episodes of sore throat	
Reducing visits to the GP for sore throat	
Improving your overall quality of life	
Improving voice	
Improving snoring	
Reducing bad breath	
Reducing financial cost to your community through NHS spending	
Reducing short term risk to taste disturbance	

Reducing risk of bleeding after treatment	
Reducing chance of altering immune system	

Thank you for taking the time to read this information letter and complete the exercise. Our research would not be possible without volunteers

Appendix H - Patient Focus group information sheet

Dear Patient,

Doctor –patient communication during medical decision-making: An observational study

Following our conversation I wanted to thank you for agreeing to participate in our focus group and attend a meeting on (Insert time and date here). You have been selected to participate in this group because you had to make a decision of which treatment to choose for recurring sore throat. The aim of the focus group will be to reach a group consensus on which 7 factors are most important when deciding which treatment to choose for the treatment of recurring sore throat.

The focus group will be audio-recorded so that we can analyse the results. All information that you provide through the following exercise and during the focus group will be kept confidentially. When you arrive you will be asked to sign a consent form to say that you agree to participate. Should you have any further questions about the study please do not hesitate to contact me on the number at the end of this letter.

Prior to our meeting I would be grateful if you could complete the following exercise and send back your answers to Nishchay.mehta.12@ucl.ac.uk.

During the focus group we will discuss why each of these factors are important for you. Then we will ask you to repeat the exercise at the end of the focus group.

Exercise

Please rank, in order of importance, which factors about a treatment and outcome are most important to you when deciding how best to treat recurring sore throat? (Order the list from 1-10, with 1 meaning most important for you in choosing a treatment, and 10 being the least important to you)

Factor	Your rank (1-14) of importance in your decision
Reducing days of sore throat	
Reducing number of episodes of sore throat	
Reducing visits to the GP for sore throat	
Improving your quality of life	
Reducing societal cost burden	
Reducing short term risk to taste disturbance (up to 6 months)	
Reducing short term risk of bleeding after treatment	
Reducing chance of altering immune system	
Reducing risk of change in voice	

Reducing halitosis	

Thank you for taking the time to read this information letter and complete the exercise. Our research would not be possible without volunteers.

Dr Nish Mehta

Nishchay.mehta.12@ucl.ac.uk

+44 20 3549 5559

evidENT

330 Grays Inn Rd, London WC1X 8DA

Appendix I – Consent for focus group

Study Number: 14/0876

CONSENT FORM FOR FOCUS GROUP PARTICIPANTS

Title of Project: Doctor –patient communication during medical decision-making: An observational study

Name of Researcher: **Dr Nish Mehta**

Please initial all boxes

1. I confirm that I have read and understand the information sheet dated 5th May 2015 Version 1.1 for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐
2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my medical care or legal rights being affected. ☐

3. I understand that my information will be kept confidentially at all times and I will not be named on any of the reports or publications that result from this study. ☐

4. I agree to having the focus group audio-recorded to allow later analysis of the discussion. ☐

5. I understand that data collected during the study may be looked at by individuals from the sponsor of the study (University College London Hospitals) to ensure the study is conducted to a high standard. I give permission for these individuals to have access to any materials resulting from the focus group ☐

6. I agree to take part in the above study. ☐

Name of Patient Date Signature

Name of Person taking consent Date Signature

Nishchay Mehta 21/05/2015

Name of Chief Investigator Date Signature

Appendix J – Semi-structured interview guide

1. House rules
    a. Toilets
    b. Fire escape
    c. Try and speak one at a time
    d. Confidential
    e. Tape recording
2. Consent
3. Introductions and Purpose of exercise
4. Exercise 1: Questions/ comments on the activity/ task
    a. Do you have any questions or comments about this exercise? Or activity
    b. How clear did you find it?
5. Exercise 2: Wording and understanding
    a. What did you understand from the word societal costs?
    b. Immunity
    c. Halitosis
6. Exercise 2: Justification for ranking - Limiting to top 7
    a. Which factors were the most important to you?
    b. Which factors were least important to you?
    c. Tell me how you went about ranking your top 3?
    d. Tell me about how you ranked the bottom 3?
    e. How did you make your choices?
7. Exercise 3: Can you add anything to this list of 10 factors? Or anything missed?
    a. Why do you want to add that
    b. Where would you rank it in your original list?

8. Exercise 4: Can each of you now please repeat the ranking exercise
    a. Did you use a different approach?
    b. Did you change any rankings?
    c. Are any of these factors similar to each other?
    d. Group on table
9. Concluding
    a. Evaluation sheet
    b. Contact if you want to hear how this goes

Preparation tasks

- Welcome by me
- Aneeka to bring expense forms
- Get name badges/stickers – Aneeka?- yes will bring
- I will arrange chairs in circles and add low lying table
- Aneeka to bring tape recorder- confirmed
- I will bring A4 papers blank and filled with marker pens
- I will bring 4 small writing pads and pens
- Aneeka to ask relevant people to complete prioritisation exercise if not already done and complete travel expense forms
- I will facilitate group
- Aneeka will take notes and facilitate if topic is being missed

Appendix K - Draft version of PARTT

Patient Unique Identifier _____ Doctor Unique Identifier _____

Investigations into patient –doctor decision making

This questionnaire is about how you and your doctor came to the decision of how best to treat your recurring sore throats (the decision making process). The questions are about which factors were important for you when making your decision.

Making Choices about Treatment

You have just seen your doctor and discussed whether to have an operation under a general anaesthetic to remove your tonsils or whether to wait and see if your symptoms of recurring sore throat resolve without an operation over time.

The next set of questions will help us as researchers understand which preferences are important to you when you made your decision about your treatment.

For each question, we ask you to decide which one was most important when you decided to have tonsillectomy or not. Please mark your answer by placing a cross on the scale shown below. If both pieces of information are equally important to you then please place your cross in the middle.

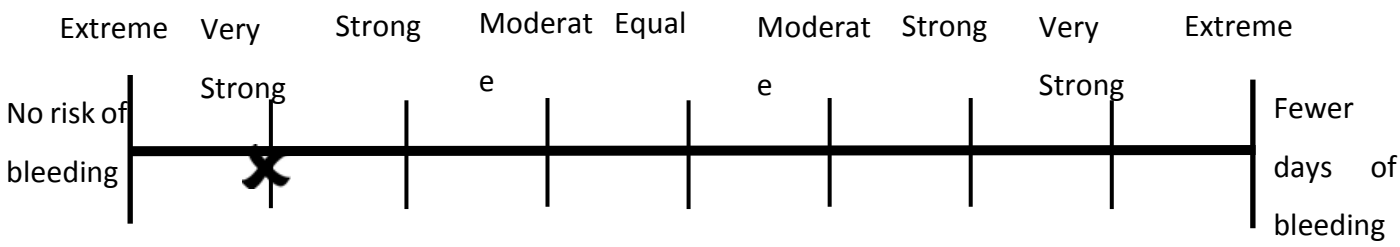
The following information may help you make your choices:

1. A tonsillectomy can mean you have up to 11 fewer days of sore throat in 6 months.
2. There is a small chance of bleeding following a tonsillectomy. Bleeding is not fatal but would require you to return to hospital for special treatment.
3. A tonsillectomy can mean you make one less visit to the GP in the first 6 months after surgery.
4. After a tonsillectomy, some patients report a large improvement in their quality of life

5. Some people who had bad breath from the recurring sore throats reported better breath following tonsillectomy.

Before you start, here is a simple example of how to answer the questions:

Please indicate your answer by placing a cross on the scale below



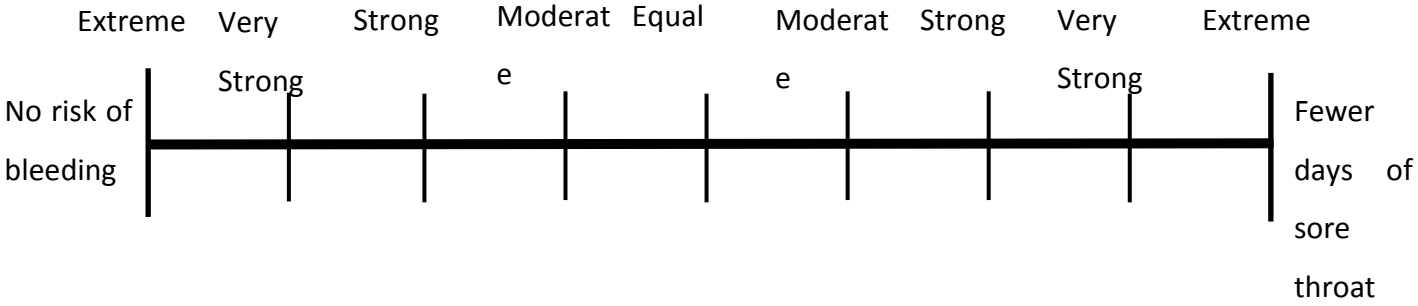
In this example, the person had a very **strong preference for No risk of bleeding compared to Fewer days of bleeding** so he/she placed a cross on "very strong" at the *'No risk of bleeding'* end of the scale.

Now please answer the questions below:

Thinking about your decision to have a tonsillectomy or not please answer the following questions in the same way.

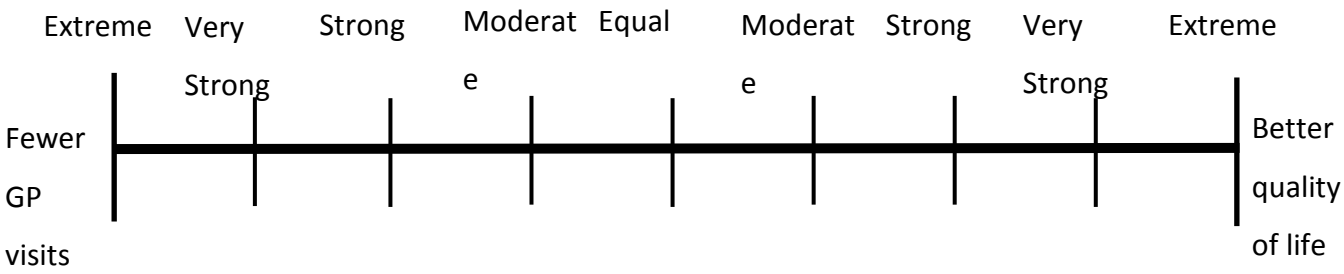
1. What is your preference: fewer days of sore throat or no risk of bleeding from a treatment?

Please indicate your answer by placing a cross on the scale below



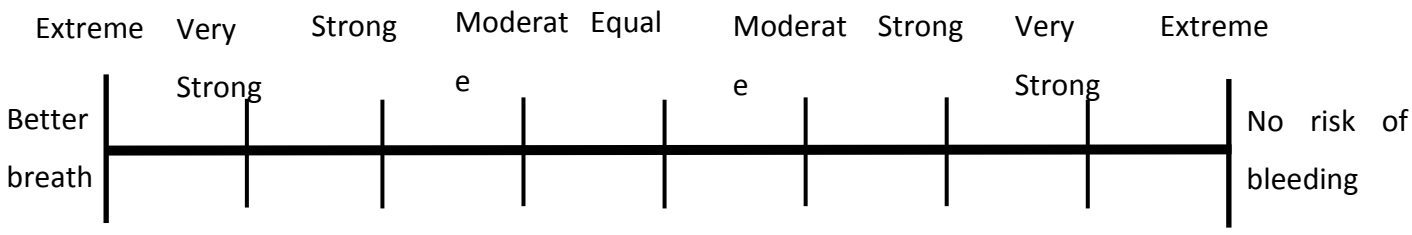
2. What is your preference: fewer visits to the GP or better quality of life?

Please indicate your answer by placing a cross on the scale below



3. What is your preference: No risk of bleeding from a treatment or having better breath?

Please indicate your answer by placing a cross on the scale below



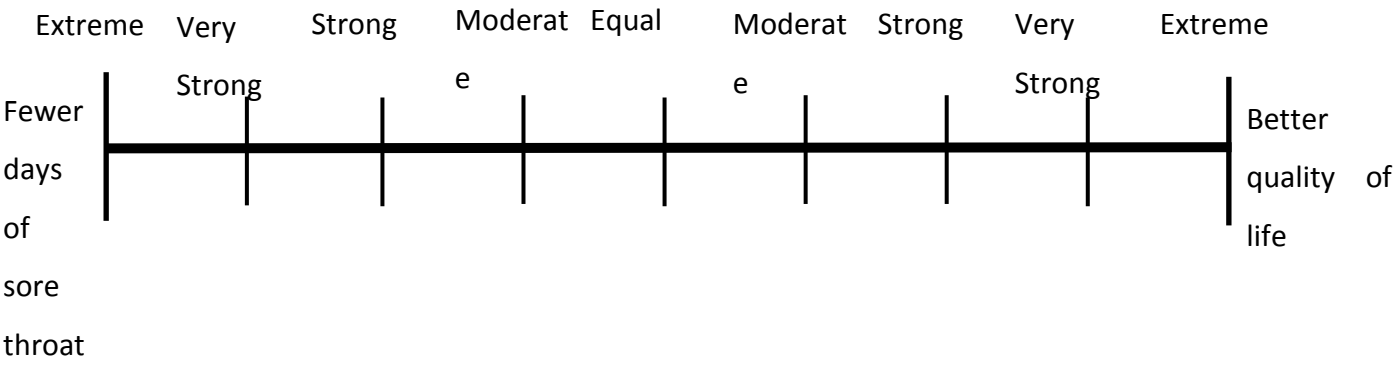
4. What is your preference: fewer days of sore throat or better quality of life?

Before you decide, please consider these two pieces of information:

A tonsillectomy can mean you have up to 11 fewer days of sore throat in 6 months.

After a tonsillectomy, some patients report a large improvement in their quality of life

Please indicate your answer by placing a cross on the scale below



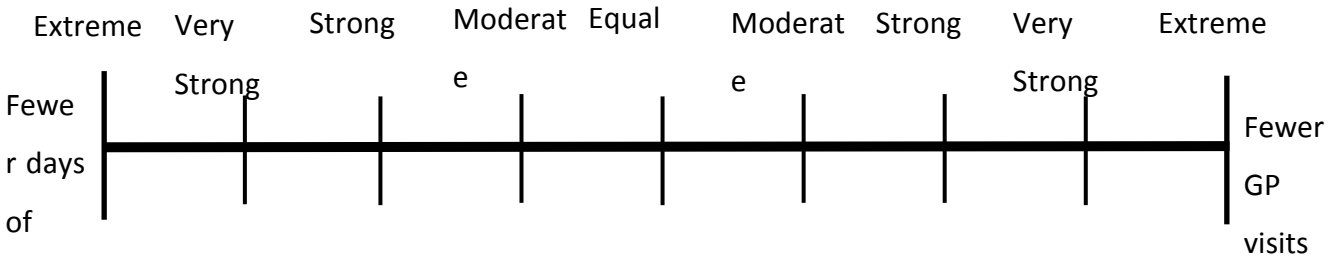
5. What is your preference: Fewer days of sore throat or fewer visits to the GP?

Before you decide, please consider these two pieces of information:

A tonsillectomy can mean you have up to 11 fewer days of sore throat in 6 months.

A tonsillectomy can mean you make one less visit to the GP in the first 6 months after surgery.

Please indicate your answer by placing a cross on the scale below



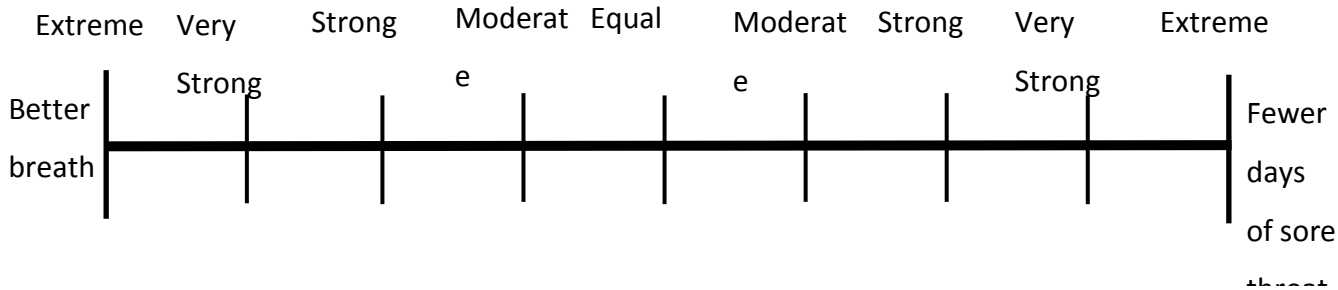
6. What is your preference: Fewer days of sore throat or having better breath?

Before you decide, please consider these two pieces of information:

A tonsillectomy can mean you have up to 11 fewer days of sore throat in 6 months

Some people who had bad breath from the recurring sore throats reported better breath following tonsillectomy

Please indicate your answer by placing a cross on the scale below



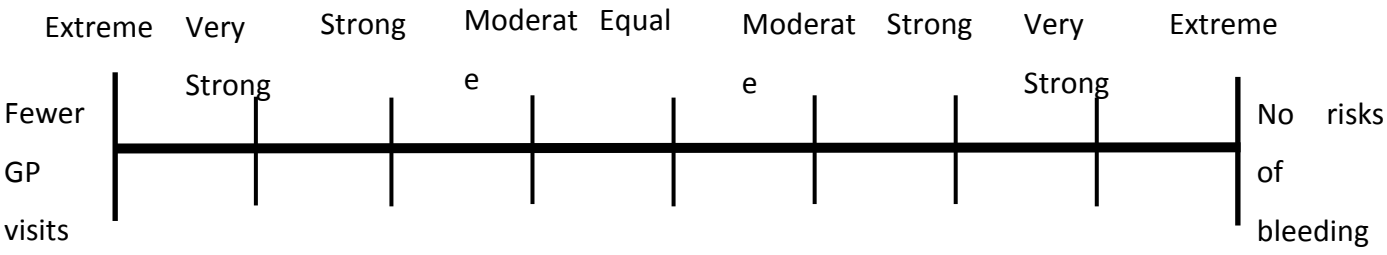
7. What is your preference: Less visits to the GP or no risk of bleeding from a treatment?

Before you decide, please consider these two pieces of information:

A tonsillectomy can mean you make one less visit to the GP in the first 6 months after surgery.

There is a small chance of bleeding following a tonsillectomy. Bleeding is not fatal but would require you to return to hospital for special medications.

Please indicate your answer by placing a cross on the scale below



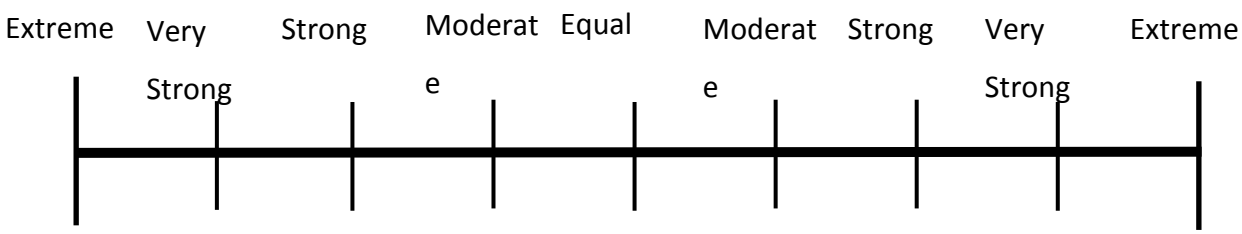
8. What is your preference: Fewer visits to the GP or having better breath?

Before you decide, please consider these two pieces of information:

A tonsillectomy can mean you make one less visit to the GP in the first 6 months after surgery.

Some people who had bad breath from the recurring sore throats reported better breath following tonsillectomy.

Please indicate your answer by placing a cross on the scale below



Fewer GP visits

Better breath

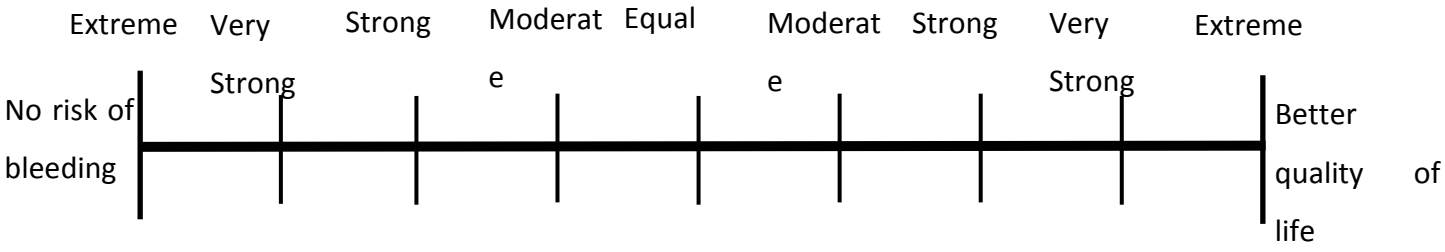
9. What is your preference: Better quality of life or No risk of bleeding?

Before you decide, please consider these two pieces of information:

After a tonsillectomy, some patients report a large improvement in their quality of life

There is a small chance of bleeding following a tonsillectomy. Bleeding is not fatal but would require you to return to hospital for special medications.

Please indicate your answer by placing a cross on the scale below



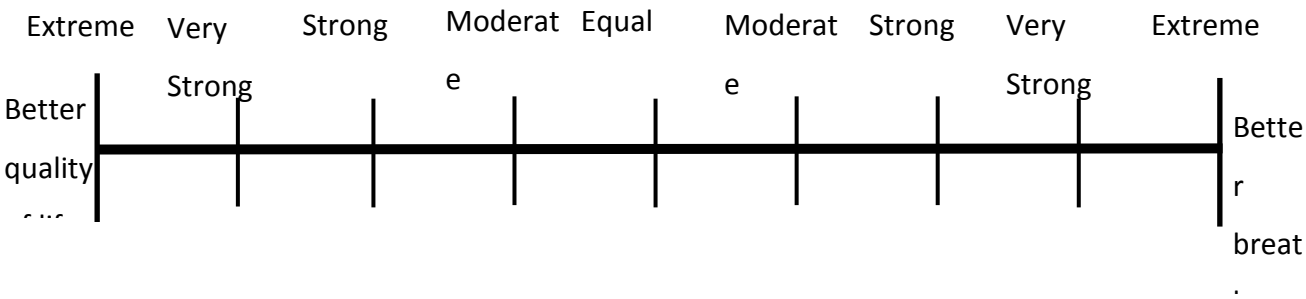
10. What is your preference: Better quality of life or better breath?

Before you decide, please consider these two pieces of information:

After a tonsillectomy, some patients report a large improvement in their quality of life (medium quality research studies)

Some people who had bad breath from the recurring sore throats reported better breath following tonsillectomy.

Please indicate your answer by placing a cross on the scale below



Thank you completing the questionnaire. Please now give the questionnaire to your researcher.

Appendix L – Final PARTT

Patient Unique Identifier _____ Doctor Unique Identifier _____

Investigations into patient –doctor decision making

This questionnaire is about how you and your doctor came to the decision of how best to treat your recurring sore throats (the decision making process). The first set of questions are about your certainty during the decision making process, the second set of questions relate to your satisfaction with the decision making process, the third set of questions are about which factors were important for you when making your decision and the final section is about your personal background.

1. Which treatment option did you and the doctor decide upon (Please tick only one)?

☐ Tonsillectomy ☐ Watch and Wait

Thinking about the decision you have just made with your doctor please consider each of the following statements and for each statement tick the box that you agree with most.

Uncertainty

	Strongly disagree	**Disagree**	**Neither agree or disagree**	**Agree**	**Strongly agree**
2. I know which options are available to me	☐	☐	☐	☐	☐

3. I know the benefits of each option	☐	☐	☐	☐	☐
4. I know the risks and side effects of each option	☐	☐	☐	☐	☐
5. I am clear about which benefits matter most to me	☐	☐	☐	☐	☐
6. I am clear about which risks and side effects matter most to me	☐	☐	☐	☐	☐
	Strongly disagree	Disagree	Neither agree or disagree	Agree	Strongly agree
7. I am clear about which is more important (the benefits or the risks and side effects)	☐	☐	☐	☐	☐
8. I have enough support from others to make a choice	☐	☐	☐	☐	☐
9. I am choosing without pressure from others	☐	☐	☐	☐	☐
10. I have enough advice to make a choice	☐	☐	☐	☐	☐
11. I am clear about the best choice for me	☐	☐	☐	☐	☐
12. I feel sure about what to choose	☐	☐	☐	☐	☐

13. This decision is easy for me to make	☐	☐	☐	☐	☐
14. I feel I have made an informed choice	☐	☐	☐	☐	☐
15. My decision shows what is important to me	☐	☐	☐	☐	☐
16. I expect to stick to my decision	☐	☐	☐	☐	☐
17. I am satisfied with my decision	☐	☐	☐	☐	☐

Satisfaction

The following questions are about how you reached your treatment decision. For each statement, please tick the box that you agree with most.

	Strongly disagree	**Disagree**	**Neither agree or disagree**	**Agree**	**Strongly agree**
18. I am satisfied that I am adequately informed about the issues important to my decision.	☐	☐	☐	☐	☐

19. The decision I made was the best decision possible for me personally.	☐	☐	☐	☐	☐
20. I am satisfied that my decision was consistent with my personal values.	☐	☐	☐	☐	☐
21. I expect to successfully carry out (or continue to carry out) the decision I made.	☐	☐	☐	☐	☐
22. I am satisfied that this was my decision to make	☐	☐	☐	☐	☐
23. I am satisfied with my decision	☐	☐	☐	☐	☐

Making Choices about Treatment

You have just seen your doctor and discussed whether to have an operation under a general anaesthetic to remove your tonsils or whether to wait and see if your symptoms of recurring sore throat resolve without an operation over time.

The next set of questions will help us as researchers understand which preferences are important to you when you made your decision about your treatment.

For each question, you will be given two pieces of information. We ask you to compare these 2 pieces of information and then decide which one was most important when you decided to have tonsillectomy or not. Please mark your answer by placing a cross on the scale shown below. **If both pieces of information are equally important to you then please place your cross in the middle.**

Before you start, here is a simple example of how to answer the questions:

Imagine you want to get from your house to the shops and you can go by either bus or bicycle.

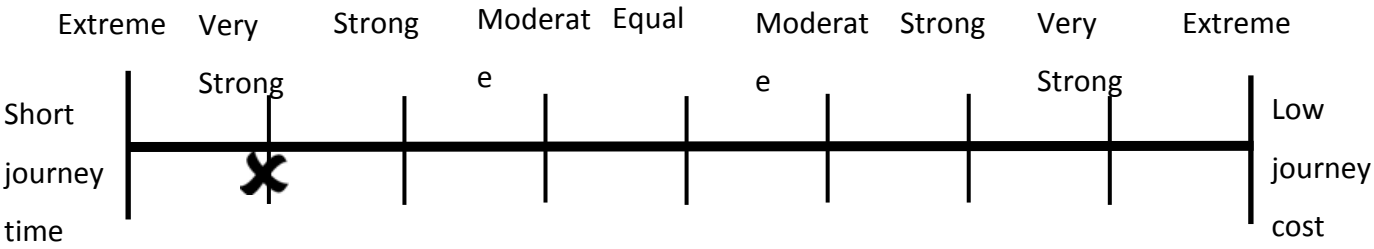
What is your preference: short journey time or low journey cost?

Before you decide, please consider these two pieces of information:

Riding a bus will save you 20 minutes in journey time.

Riding a bicycle will be £4 cheaper.

Please indicate your answer by placing a cross on the scale below



In this example, the person had a very strong preference for short journey time compared with low journey cost so he/she placed a cross on “very strong” at the short journey time end of the scale. If the person felt that low journey cost was as important to them as short journey time, they would have put their cross in at the Equal marker.

Thinking about your decision to have a tonsillectomy or not please answer the following questions in the same way.

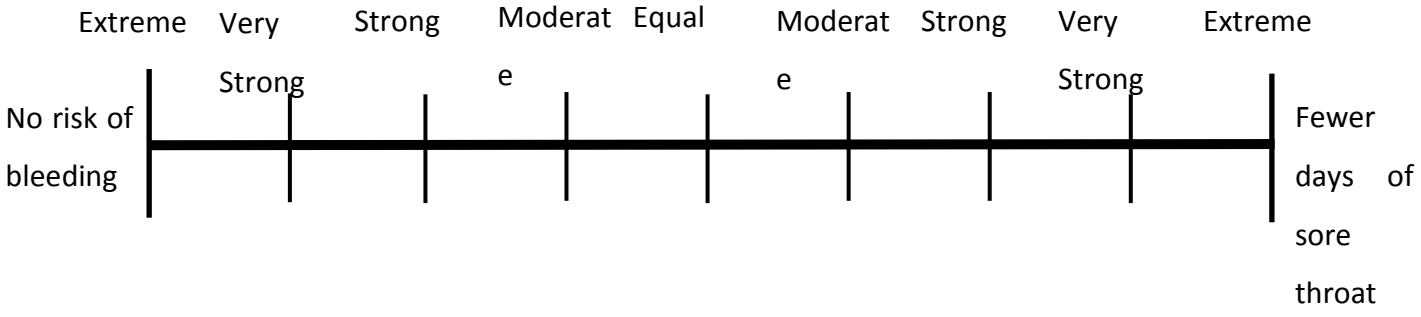
24. What is your preference: fewer days of sore throat or no risk of bleeding from a treatment?

Before you decide, please consider these two pieces of information:

A tonsillectomy can mean you have up to 11 fewer days of sore throat in 6 months.

There is a small chance of bleeding following a tonsillectomy. Bleeding is not fatal but would require you to return to hospital for special treatment.

Please indicate your answer by placing a cross on the scale below



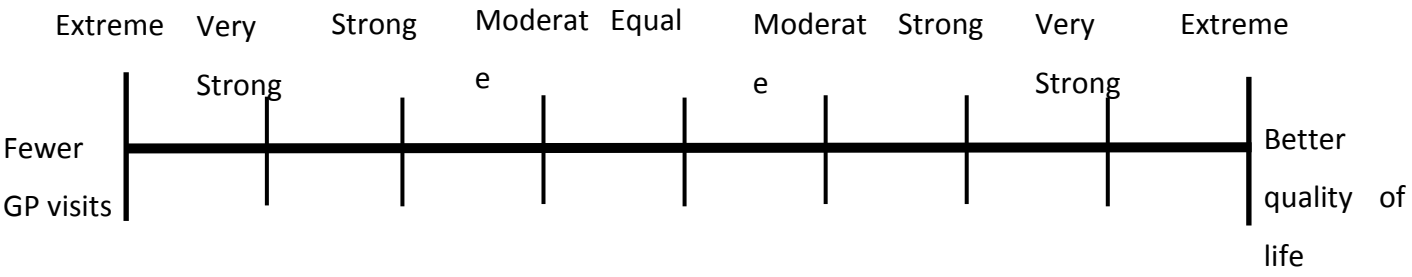
25. What is your preference: fewer visits to the GP or better quality of life?

Before you decide, please consider these two pieces of information:

A tonsillectomy can mean you make one less visit to the GP in the first 6 months after surgery.

After a tonsillectomy, some patients report a large improvement in their quality of life

Please indicate your answer by placing a cross on the scale below



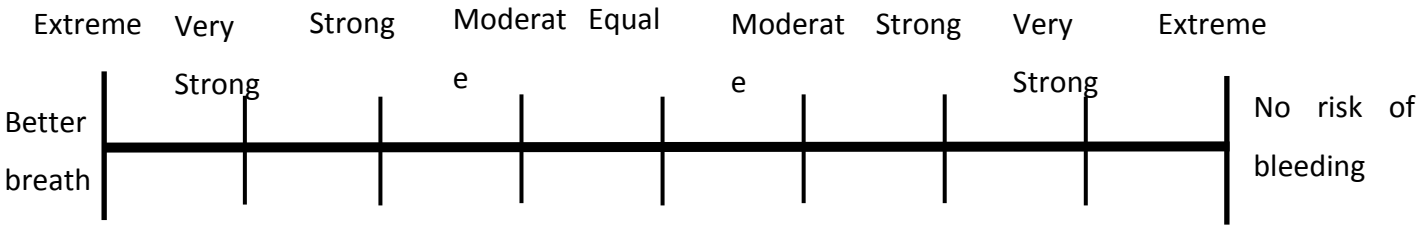
26. What is your preference: No risk of bleeding from a treatment or having better breath?

Before you decide, please consider these two pieces of information:

Some people who had bad breath from the recurring sore throats reported better breath following tonsillectomy.

There is a small chance of bleeding following a tonsillectomy. Bleeding is not fatal but would require you to return to hospital for special treatment.

Please indicate your answer by placing a cross on the scale below



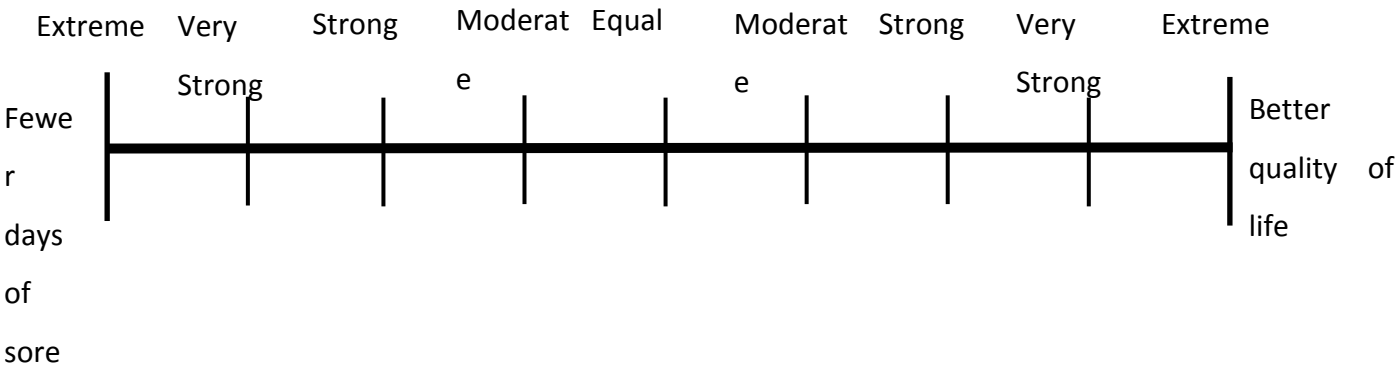
27. What is your preference: fewer days of sore throat or better quality of life?

Before you decide, please consider these two pieces of information:

A tonsillectomy can mean you have up to 11 fewer days of sore throat in 6 months.

After a tonsillectomy, some patients report a large improvement in their quality of life

Please indicate your answer by placing a cross on the scale below



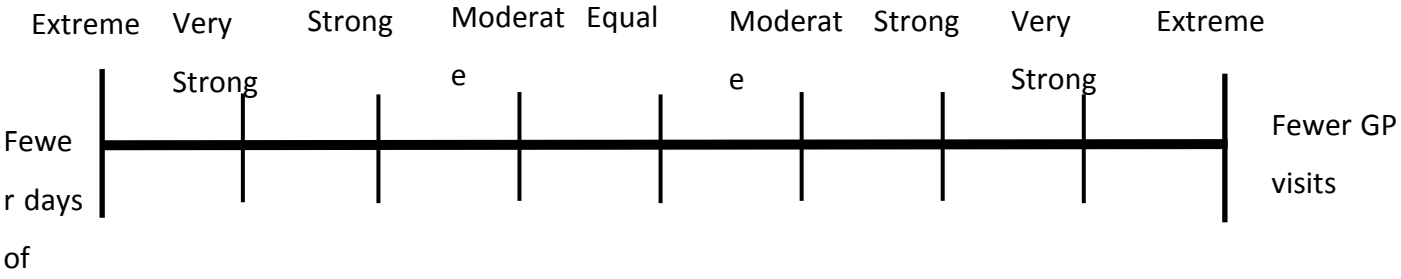
28. What is your preference: Fewer days of sore throat or fewer visits to the GP?

Before you decide, please consider these two pieces of information:

A tonsillectomy can mean you have up to 11 fewer days of sore throat in 6 months.

A tonsillectomy can mean you make one less visit to the GP in the first 6 months after surgery.

Please indicate your answer by placing a cross on the scale below



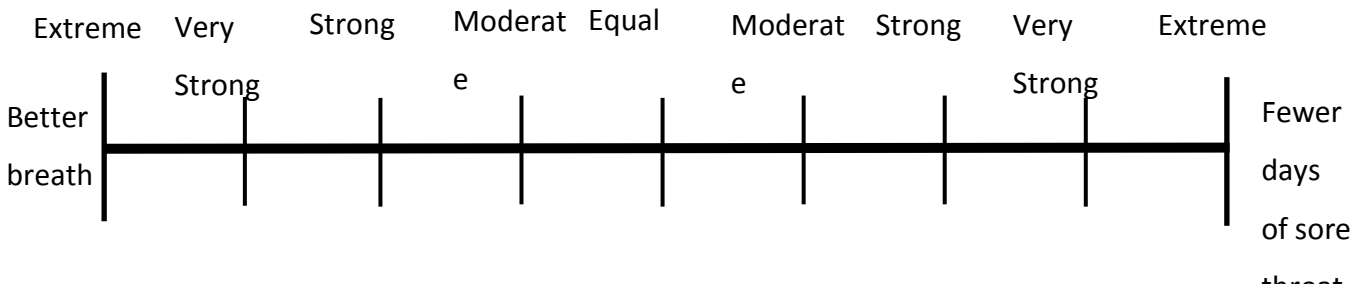
29. What is your preference: Fewer days of sore throat or having better breath?

Before you decide, please consider these two pieces of information:

A tonsillectomy can mean you have up to 11 fewer days of sore throat in 6 months

Some people who had bad breath from the recurring sore throats reported better breath following tonsillectomy

Please indicate your answer by placing a cross on the scale below



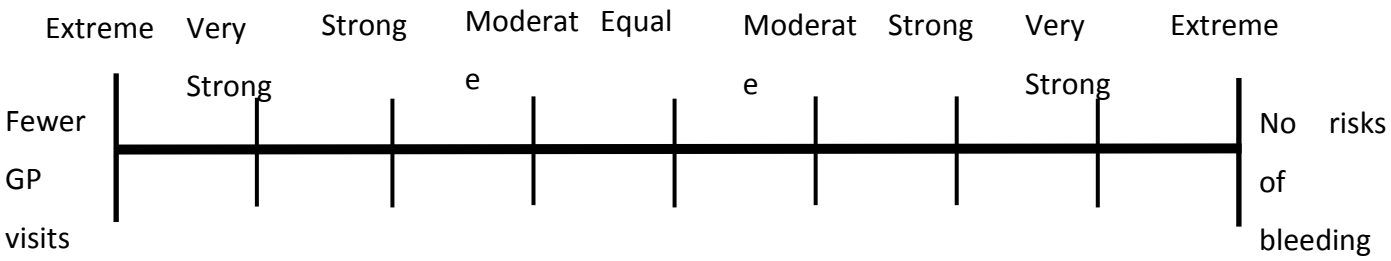
30. What is your preference: Less visits to the GP or no risk of bleeding from a treatment?

Before you decide, please consider these two pieces of information:

A tonsillectomy can mean you make one less visit to the GP in the first 6 months after surgery.

There is a small chance of bleeding following a tonsillectomy. Bleeding is not fatal but would require you to return to hospital for special medications.

Please indicate your answer by placing a cross on the scale below



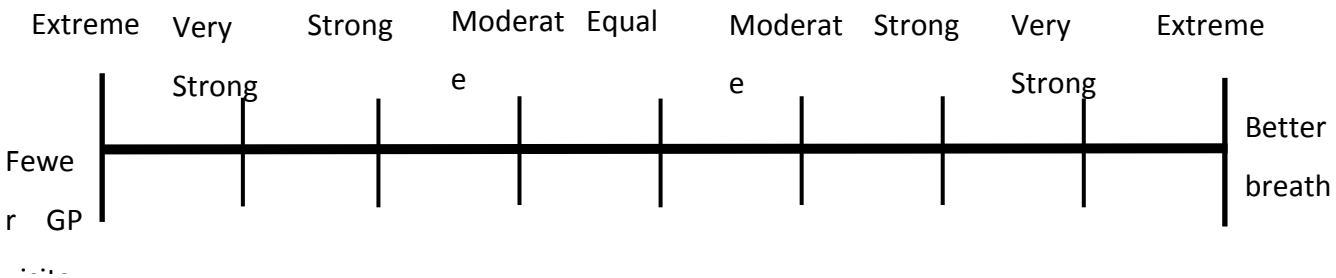
31. What is your preference: Fewer visits to the GP or having better breath?

Before you decide, please consider these two pieces of information:

A tonsillectomy can mean you make one less visit to the GP in the first 6 months after surgery.

Some people who had bad breath from the recurring sore throats reported better breath following tonsillectomy.

Please indicate your answer by placing a cross on the scale below



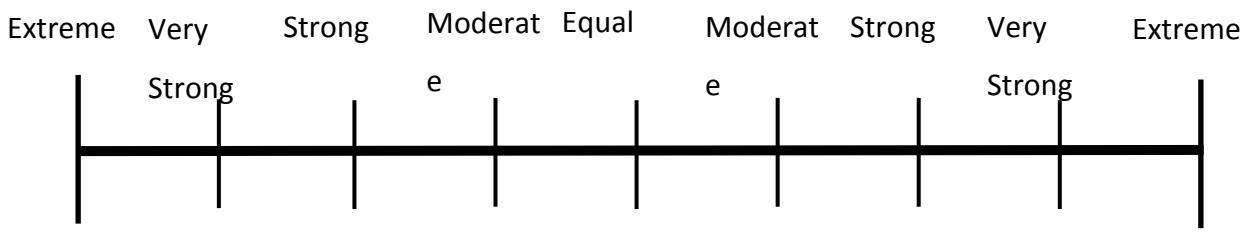
32. What is your preference: Better quality of life or No risk of bleeding?

Before you decide, please consider these two pieces of information:

After a tonsillectomy, some patients report a large improvement in their quality of life

There is a small chance of bleeding following a tonsillectomy. Bleeding is not fatal but would require you to return to hospital for special medications.

Please indicate your answer by placing a cross on the scale below



No risk of

Better quality of life

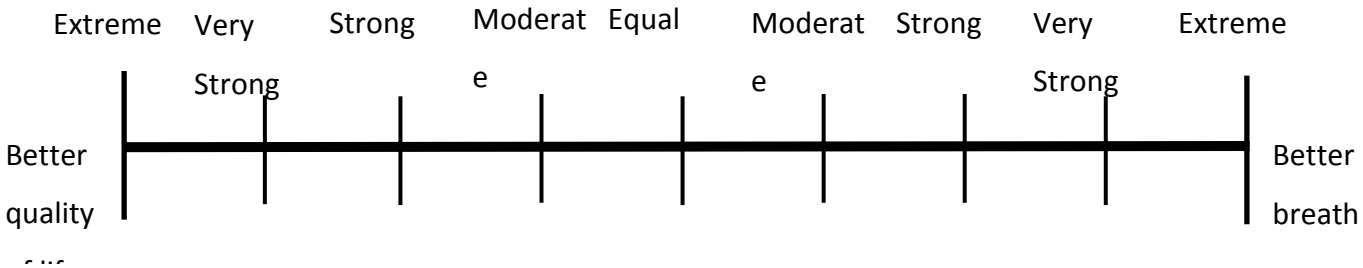
33. What is your preference: Better quality of life or better breath?

Before you decide, please consider these two pieces of information:

After a tonsillectomy, some patients report a large improvement in their quality of life (medium quality research studies)

Some people who had bad breath from the recurring sore throats reported better breath following tonsillectomy.

Please indicate your answer by placing a cross on the scale below



Socio-demographics

Now we would like to ask you some background questions. For each question, tick one box only.

34. How old are you?

☐ 15-19 ☐ 20-29 ☐ 30-39 ☐ 40-49 ☐ 50-59 ☐ 60-69

35. What is your sex?

☐ Male ☐ Female

36. Which of the following best describes your ethnic group?

☐ White

☐ Mixed / Multiple ethnic groups

☐ Asian / Asian British

☐ Black / African / Caribbean / Black British

☐ Any other ethnic group, please describe _____

37. Current occupation (please describe previous if unemployed / retired)?

38. In your opinion, how was the timing of referral from your GP to our ENT services?

☐ Prompt ☐ Appropriate ☐ Delayed

39. How long have you been suffering with severe sore throats?

_____Years

40. How many episodes of sores throat have you had in the last 12 months?

☐ 0-3 ☐ 4-5 ☐ 6-7 ☐ More than 8

41. How many days off work/education have you had to take in the last 12 months because of sore throats?

☐ 0-5 ☐ 6-10 ☐ 11-15 ☐ More than 16

Thank you completing the questionnaire. Please now give the questionnaire to your researcher.

Appendix M - Detailed Doctor Information Sheet

Detailed Doctor Information Sheet

Doctor –patient communication during medical decision-making: An observational study

Protocol Reference Number: 14/0876

We would like to invite you to take part in our research study. Before you decide we would like you to understand why the research is being done and what it would involve for you. One of our team will give a small talk on this study and will answer any questions you may have.

Part 1 tells you the purpose of this study and what will happen to you if you take part. Part 2 gives you more detailed information about the conduct of the study.

Ask us if there is anything that is not clear. Take time to decide whether or not you wish to take part.

Part1

What is the purpose of the study?

We want patients to be more involved in decisions that will affect their treatment. Currently we have little understanding of how decisions about medical treatments are being made. The aim of this study is to get a better understanding of how decisions between doctors and patients are made when there is no 'best treatment' option available. This information will then be used to help doctors and patients communicate better when making decisions.

Why am I being asked to participate in this study?

We will be recruiting 150 adults who have been considered for a tonsillectomy to treat recurring sore throats across 10 hospitals in England (15 patients/hospital) over 6 months. We will be asking them questions about their level of certainty when the decision was made and their overall satisfaction with it. However, we also want to know the same information about the doctors who saw them in clinic so that we can build a clearer picture of decision-

making process. We are inviting you to participate, as you are likely to see one of these patients during a routine outpatient clinic.

Do I have to take part?

No. Taking part is completely voluntary. We will describe the study during the talk we give to your department. If you agree to take part, we will then ask you to sign a consent form. You are free to withdraw at any time without giving a reason.

What will happen to me if I take part?

If you agree to participate in the study you will be asked to complete a short 5 - minute questionnaire on your personal values and sociodemographic details. Subsequently, you will be asked to complete a 2-minute paper questionnaire for every patient you see that has also been recruited into the study. The questions will relate to decisions made during the consultation.

What are the possible benefits of taking part?

A decision of which treatment to take can be difficult when the options aren't clear. We want to understand how doctors and patients make these decisions, so that we can make information more clear.

Will my taking part in the study be kept confidential?

Yes. We will follow ethical and legal practice and all information about you will be handled in confidence. The details are included in Part 2.

Contact Details

Your Local study co-ordinator

Dr Shilpa Ojha shilpa.ojha@uclh.nhs.uk

Chief Investigator

Dr Nish Mehta Nishchay.mehta.12@ucl.ac.uk

This completes Part 1 of the Information Sheet.

If the information in Part 1 has interested you and you are considering participation, please read the additional information in Part 2 before making any decision.

Part 2

Will my data be kept safe?

If you consent to take part in this study the information you provided will be kept confidentially at all times and will be stored at UCL in accordance with the Data Protection Act (1998). You will be allocated a code number, which will be used instead of your name to identify you on all study forms.

Only the project team will have access to the information you provide. The Sponsor (UCLH) may also request access to your information to ensure the study is being carried out correctly. By signing the consent form you agree to this access by the sponsor for the current study. At the end of the study your data will be securely archived for 5 years at UCL.

What will happen to the results of the research study?

The results of the study will be published in a medical journal and be presented at a scientific conference. They will also be written up as part of the PhD thesis of the Chief Investigator. No participants will be identified by name in any of the reports or publications. Should you wish to see the results, or the publications, please ask your local study co-ordinator (Dr Shilpa Ojha).

Who has reviewed the study?

This study has been reviewed and given favourable opinion by Nottingham Research Ethics Committee.

Complaints

If you have a concern about any aspect of this study, you should ask to speak to your local study co-ordinator researcher who will do their best to answer your questions (Dr Shilpa Ojha details above). If you remain unhappy and wish to complain formally, you can contact the Chief Investigator (Mr Nish Mehta Nishchay.mehta.12@ucl.ac.uk).

Further information and contact details

If you have any questions about the study, please speak to your local study co-ordinator (Dr Shilpa Ojha, details above), who will be able to provide you with up to date information about the procedure involved. If you require any further information or have any concerns while taking part in the study please contact:

Dr Nishchay Mehta +44 20 3549 5559

If you decide you would like to take part then please read and sign the consent form. You will be given a copy of the information sheet and the consent form will be filed with the study records.

You can have more time to think this over if you are at all unsure.

Thank you for taking the time to read this information sheet and to consider this study.

Study Collaborators

UCL
wellcome trust
UCLH NHS
NHS Foundation Trust

Appendix N - Doctor Consent Form

Centre Number: 1

Study Number: 14/0876

Participant Identification Number for this study:

CONSENT FORM FOR DOCTORS

Title of Project: Doctor –patient communication during medical decision-making: An observational study

Name of Researcher: **Dr Nish Mehta**

Please initial all boxes

7. I confirm that I have read and understand the information sheet dated 5th May 2015 Version 1.1 for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐
8. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason legal rights being affected. ☐
9. I understand that the data collected during the study may be looked at by individuals from the sponsor of the trial (University College London Hospitals) to ensure the study is conducted to a high standard. I give permission for these individuals to have access to my details and questionnaire. ☐
10. I understand that my information will be kept confidentially at all times and I will not be named on any of the reports or publications that result from this study. ☐

☐

11. I agree to take part in the above study.

Name of Participant	Date	Signature

Name of Person taking consent	Date	Signature

Nishchay Mehta	21/05/2015	

Name of Chief Investigator	Date	Signature

Appendix O - Patient Invitation Letter

Dear Patient,

Doctor –patient communication during medical decision-making: An observational study

We are currently undertaking a study to see how patients and doctors make decisions related to medical problems.

The goal of this study is to help doctors and patients communicate better to make easier and safer decisions in the future. You may be asked to participate in this study when you have completed your consultation with the ENT surgeon.

Why am I being sent this letter?

We will be recruiting 150 adults who have recurring sore throats throughout hospitals in England. The referral letter your GP has sent to our department suggests you may be eligible for our study. At the end of your consultation, if your ENT doctor agrees that you are eligible for this study you will be asked if you want to participate.

Do I have to take part?

No. Taking part is completely voluntary. If your ENT doctor agrees that you are eligible for this study it will be up to you to decide whether you want to join or not. If you agree to take part, we will then ask you to sign a consent form. You are free to withdraw at any time without giving a reason. This would not affect the standard of care you receive.

Why should I take part?

A decision of which treatment to take can be difficult when the options aren't clear. We want to understand how doctors and patients make these decisions so that we can make information more clear.

What will it involve?

If you agree to participate in the study you will be asked to complete a 10-minute paper questionnaire. The questions will be related to the decision you made together with your doctor and your satisfaction with it. In addition you will be asked a few questions about your age, gender, education, occupation and ethnicity. This will help us tailor our planned improvements to different patient groups. Your treatment will not be affected by how you fill in your questionnaire. All the information we collect will be kept confidentially and any completed questionnaires will not have your name on them.

What do I do now?

Nothing. When you go to your outpatients' appointment someone from our research team may meet and talk to you about this study in more detail. If your ENT doctor feels you are appropriate they will ask you if you want to take part.

Where will the information go?

The information you provide will be kept confidentially. Your name will be separated from your questionnaire. Your name and questionnaire will be linked by a secure code number.

What will you do with the results?

The results of the study will be published in a medical journal and be presented at a scientific conference. They will also be written up as part of the PhD thesis of the Chief Investigator. No participants will be identified by name in any of the reports or publications. Should you wish to see the results, or the publications, please ask the research team at the hospital for further details.

Who do I contact if I want more information or have concerns?

Please contact the Dr Nish Mehta below should you want further information.

Dr Nish Mehta

Nishchay.mehta.12@ucl.ac.uk

evidENT

330 Grays Inn Rd, London WC1X 8DA

Thank you for taking the time to read this information letter, whether or not you choose to participate in the study. Our research would not be possible without volunteers.

Appendix P – Detailed Patient Information Sheet

Detailed Patient Information Sheet

Doctor –patient communication during medical decision-making: An observational study

Protocol Reference Number: 14/0876

We may invite you to take part in our research study depending on which treatment you are offered. If you are invited to participate we would like you to understand why the research is being done and what it would involve for you. One of our team can go through the information sheet with you and answer any questions you have. We'd suggest this should take about 5 minutes. Talk to others about the study if you wish.

Part 1 tells you the purpose of this study and what will happen to you if you take part. Part 2 gives you more detailed information about the conduct of the study.

Ask us if there is anything that is not clear. Take time to decide whether or not you wish to take part.

Part1

What is the purpose of the study?

We want patients to be more involved in decisions that will affect their treatment. Currently we have little understanding of how decisions about medical treatments are being made. The aim of this study is to get a better understanding of how decisions between doctors and

patients are made when there is no 'best treatment' option available. This information will then be used to help doctors and patients communicate better.

Why may I be asked to participate in this study?

We will be recruiting 150 adults who have recurring sore throats across 10 hospitals in England. The referral letter your GP has sent suggested you may be eligible for our study. After your consultation if the ENT doctor agrees that you are eligible for this study you will be formally asked if you want to participate.

Do I have to take part?

No. Taking part is entirely voluntary. If your doctor agrees that you are eligible for this study it will be up to you to decide whether or not you should participate. If you agree to take part, we will then ask you to sign a consent form. You are free to withdraw at any time without giving a reason. Whether or not you choose to participate the standard of care you receive will not be affected.

What will happen to me if I take part?

If you agree to participate in the study you will be asked to complete a 10-minute paper questionnaire. The questions will relate to decisions made during your appointment. In addition you will be asked a few questions about your age, gender, education, occupation and ethnicity. This will help us direct our planned improvements to different patient groups. All

the information we collect will be kept confidentially and will not affect the treatment you receive from your ENT team.

What are the possible benefits of taking part?

A decision of which treatment to take can be difficult when the options aren't clear. We want to understand how doctors and patients make these decisions so that we can make information more clear.

We cannot promise the study will help you personally, but the information we get from this study should help improve communication between patients and doctors throughout the country.

What if there is a problem?

Any complaint about the way you have been dealt with during the study will be addressed. The detailed information concerning this is given in Part 2 of this information sheet. If you have any concerns or complaints you should contact the local member of the research team in the first instance (details below).

Will my taking part in the study be kept confidential?

Yes. We will follow ethical and legal practice and all information about you will be handled in confidence. The details are included in Part 2.

Contact Details

Local Study Co-ordinator

Dr. Shilpa Ojha shilpa.ojha@uclh.nhs.uk

Chief Investigator

Dr Nishchay Mehta Nishchay.mehta.12@ucl.ac.uk

This completes Part 1 of the Information Sheet.

If the information in Part 1 has interested you and you are considering participation, please read the additional information in Part 2 before making any decision.

Part 2

Will my data be kept safely?

If you consent to take part in this study, the information you provided will be kept confidentially at all times and will be stored at your hospital in accordance with the Data Protection Act (1998). Your name will not be passed to anyone else outside the research team. You will be allocated a code number, which will be used to identify you on all study forms.

Only the project team will have access to the information you provide. The sponsor (UCLH) may also request access to your information to ensure the study is being carried out correctly. By signing the consent form you agree to this access by the sponsor for the current study. At the end of the study your data will be securely archived for 5 years at UCL.

Will my GP be informed of my involvement?

Your GP will not be informed of your participation in this study as we are not changing the care you receive. Although your hospital doctor will be aware of your participation he/she will not have access to your questionnaire.

What will happen to the results of the research study?

The results of the study will be published in a medical journal and be presented at a scientific conference. They will also be written up as part of the PhD thesis of the Chief Investigator. No participants will be identified by name in any of the reports or publications. Should you wish to see the results, or the publications, please ask your researcher (shilpa.ojha@uclh.nhs.uk).

Who is organising and funding the research?

The research is being organised by the ENT research doctors at UCL hospital in collaboration with the ENT doctors at your local hospital. The research is funded by the Wellcome Trust.

Who has reviewed the study?

All research in the NHS is looked at by independent group of people, called a Research Ethics Committee, to protect your interests. This study has been reviewed and given favourable opinion by Nottingham Research Ethics Committee.

Complaints

If you have a concern about any aspect of this study you should ask to speak to the researcher (person who gave you this information sheet). They will do their best to answer your questions. If you remain unhappy and wish to complain formally, you can contact your PALS team (020 3447 3042).

Further information and contact details

You are encouraged to ask any questions you wish, before, during or after your treatment. If you have any questions about the study, please speak to the researcher (person who gave you this information sheet), who will be able to provide you with up to date information about the procedure involved. If you wish to read the research on which this study is based, please ask your researcher. If you require any further information or have any concerns while taking part in the study please contact one of the following people:

Local Study Co-ordinator

Dr. Shilpa Ojha shilpa.ojha@uclh.nhs.uk

Chief Investigator

Dr Nishchay Mehta Nishchay.mehta.12@ucl.ac.uk

If you decide you would like to take part then please read and sign the consent form. You will be given a copy of this information sheet to keep. The consent form will be filed with the study records and a copy may be sent to the Research Sponsor.

You can have more time to think this over if you are at all unsure.

Thank you for taking the time to read this information sheet and to consider this study.

Appendix Q – Patient Consent Form

Centre Number: 1

Study Number: 14/0876

Participant Identification Number for this trial:

CONSENT FORM FOR PATIENTS

Title of Project: Doctor –patient communication during medical decision-making: An observational study

Name of Researcher: **Dr Nish Mehta**

Please initial all boxes

12. I confirm that I have read and understand the information sheet dated 5th May 2015 version 1.1 for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐
13. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my medical care or legal rights being affected. ☐
14. I understand that the data collected during the study may be looked at by individuals from the sponsor of the trial (University College London Hospitals) to ensure the study is conducted to a high standard. I give permission for these individuals to have access to my completed forms. ☐
15. I understand that my information will be kept confidentially at all times and I will not be named on any of the reports or publications that result from this study. ☐

16. I agree to take part in the above study. ☐

Name of Patient Date Signature

Name of Person taking consent Date Signature

Nishchay Mehta 21.05.2015

Chief Investigator Date Signature

Appendix R – Doctor pre-consultation questionnaire

Doctor Unique Identifier _____

Doctor's preferences of treatment factors when deciding how to manage recurrent tonsillitis

The first set of questions are about your demographics and surgical experience. The following set of questions are meant to determine which factors you would consider important when choosing between tonsillectomy and watchful waiting if you were a patient with recurrent tonsillitis. Please tick only box for each question.

1. How old are you?

☐ 20-29 ☐ 30-39 ☐ 40-49 ☐ 50-59 ☐ 60-69

2. What is your sex?

☐ Male ☐ Female

3. Which of the following best describes your ethnic group?

☐ White

☐ Mixed / Multiple ethnic groups

☐ Asian / Asian British

☐ Black / African / Caribbean / Black British

☐ Any other ethnic group, please describe _____

4. What is your Clinical Grade?

☐ Consultant ☐ Registrar ☐ Staff Grade

☐ Associate Specialist ☐ Core Trainee

5. How many years have you been in ENT practice since leaving medical school?

☐ <1 ☐ 1 – 5 ☐ 6 – 10 ☐ 11-20 ☐ >20

Making Choices about Treatment

You have just seen your doctor and discussed whether to have an operation under a general anaesthetic to remove your tonsils or whether to wait and see if your symptoms of recurring sore throat resolve without an operation over time.

The next set of questions will help us as researchers understand which preferences are important to you when you made your decision about your treatment.

For each question, you will be given two pieces of information. We ask you to compare these 2 pieces of information and then decide which one was most important when you decided to have tonsillectomy or not. Please mark your answer by placing a cross on the scale shown below. **If both pieces of information are equally important to you then please place your cross in the middle.**

Before you start, here is a simple example of how to answer the questions:

Imagine you want to get from your house to the shops and you can go by either bus or bicycle.

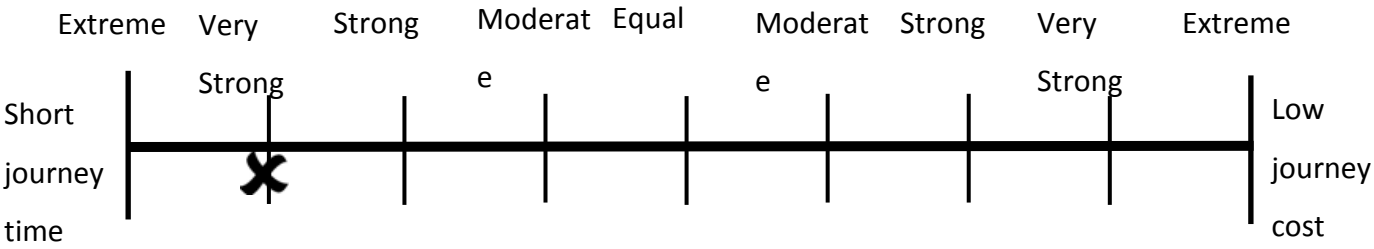
What is your preference: short journey time or low journey cost?

Before you decide, please consider these two pieces of information:

Riding a bus will save you 20 minutes in journey time.

Riding a bicycle will be £4 cheaper.

Please indicate your answer by placing a cross on the scale below



In this example, the person had a very strong preference for short journey time compared with low journey cost so he/she placed a cross on “very strong” at the short journey time end of the scale. If the person felt that low journey cost was as important to them as short journey time, they would have put their cross in at the Equal marker.

Thinking about your decision to have a tonsillectomy or not please answer the following questions in the same way.

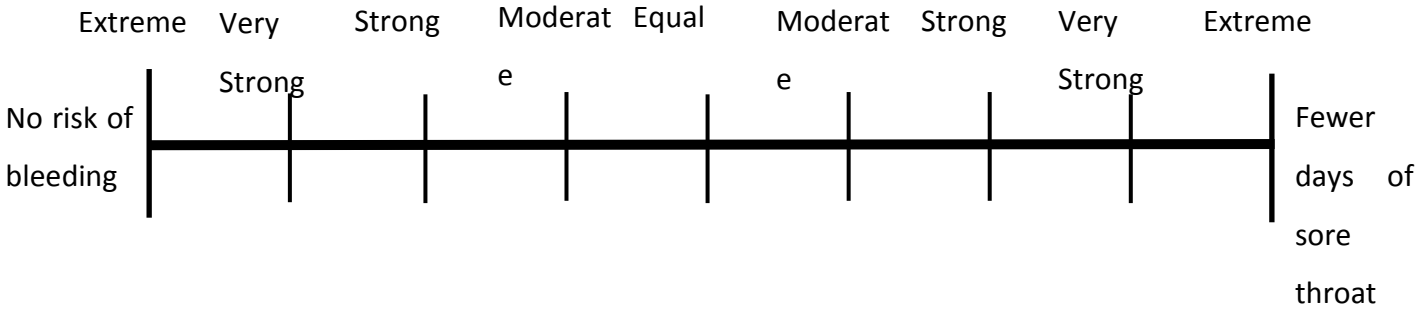
24. What is your preference: fewer days of sore throat or no risk of bleeding from a treatment?

Before you decide, please consider these two pieces of information:

A tonsillectomy can mean you have up to 11 fewer days of sore throat in 6 months.

There is a small chance of bleeding following a tonsillectomy. Bleeding is not fatal but would require you to return to hospital for special treatment.

Please indicate your answer by placing a cross on the scale below



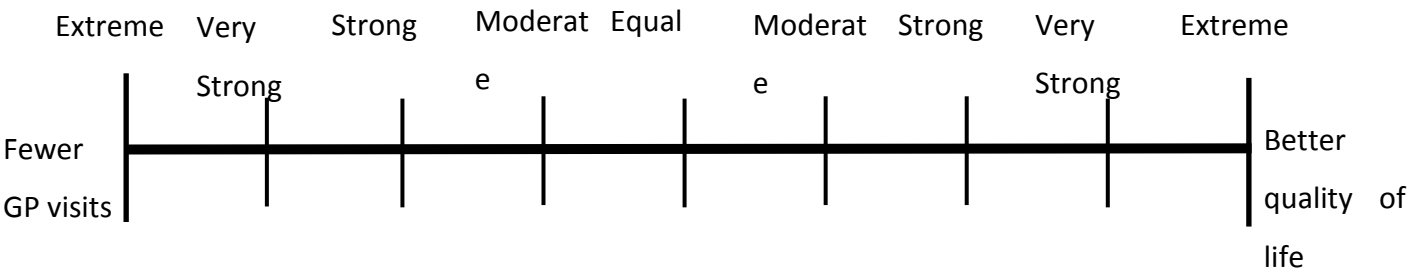
25. What is your preference: fewer visits to the GP or better quality of life?

Before you decide, please consider these two pieces of information:

A tonsillectomy can mean you make one less visit to the GP in the first 6 months after surgery.

After a tonsillectomy, some patients report a large improvement in their quality of life

Please indicate your answer by placing a cross on the scale below



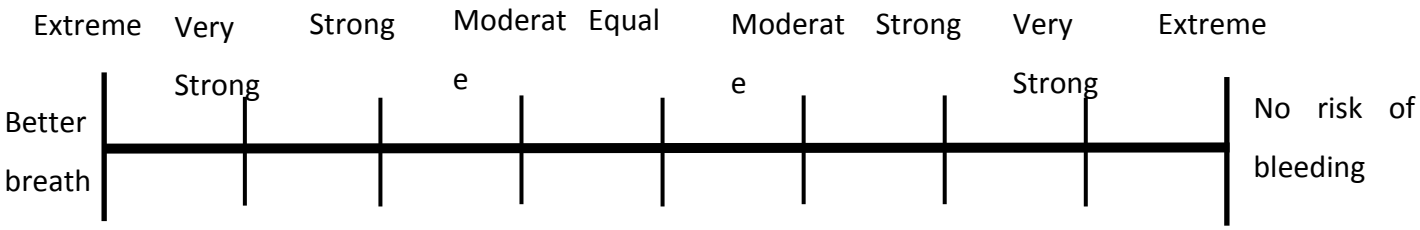
<u>26.</u> What is your preference: No risk of bleeding from a treatment or having better breath?

Before you decide, please consider these two pieces of information:

Some people who had bad breath from the recurring sore throats reported better breath following tonsillectomy.

There is a small chance of bleeding following a tonsillectomy. Bleeding is not fatal but would require you to return to hospital for special treatment.

Please indicate your answer by placing a cross on the scale below



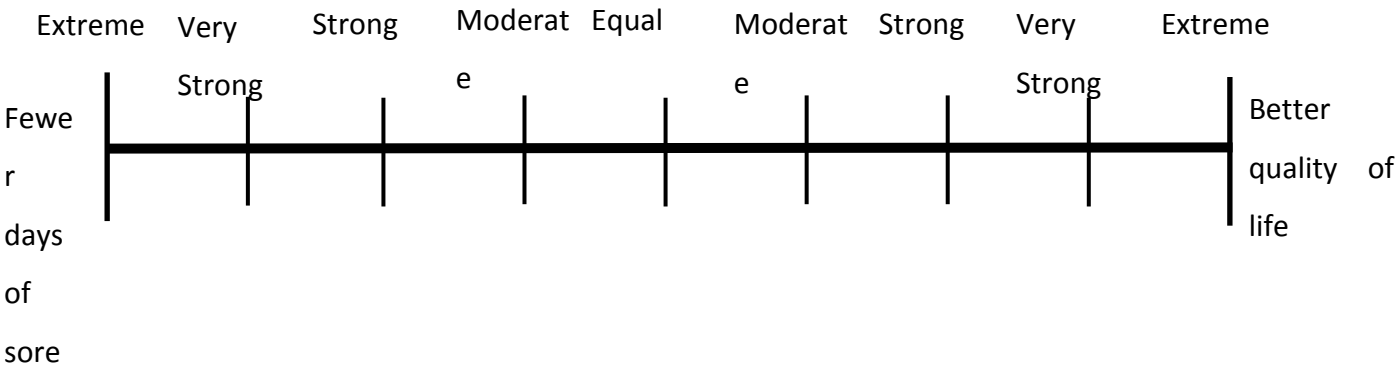
27. What is your preference: fewer days of sore throat or better quality of life?

Before you decide, please consider these two pieces of information:

A tonsillectomy can mean you have up to 11 fewer days of sore throat in 6 months.

After a tonsillectomy, some patients report a large improvement in their quality of life

Please indicate your answer by placing a cross on the scale below



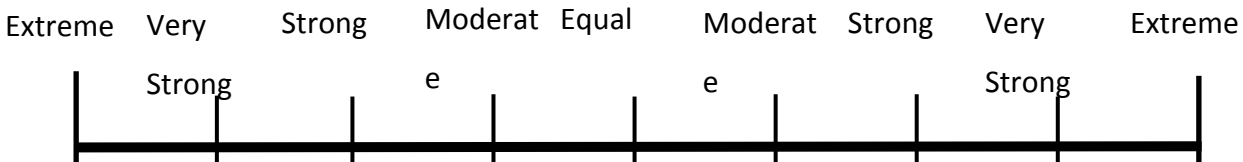
28. What is your preference: Fewer days of sore throat or fewer visits to the GP?

Before you decide, please consider these two pieces of information:

A tonsillectomy can mean you have up to 11 fewer days of sore throat in 6 months.

A tonsillectomy can mean you make one less visit to the GP in the first 6 months after surgery.

Please indicate your answer by placing a cross on the scale below



Fewe
r days
of

Fewer GP
visits

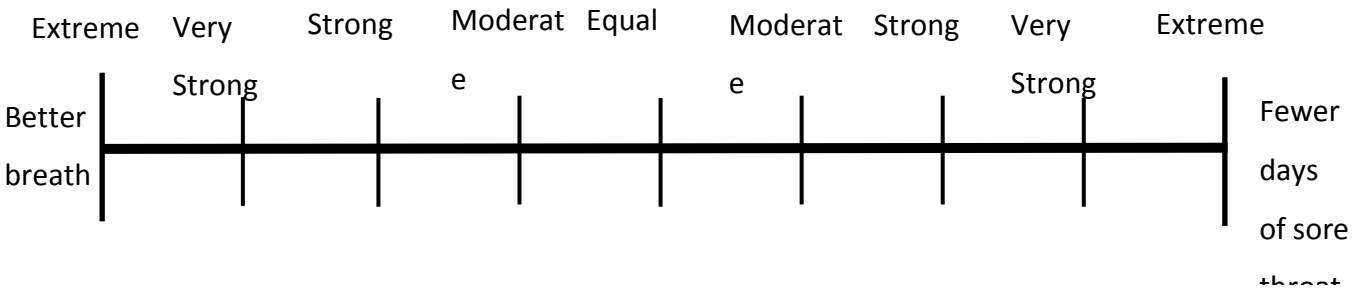
29. What is your preference: Fewer days of sore throat or having better breath?

Before you decide, please consider these two pieces of information:

A tonsillectomy can mean you have up to 11 fewer days of sore throat in 6 months

Some people who had bad breath from the recurring sore throats reported better breath following tonsillectomy

Please indicate your answer by placing a cross on the scale below



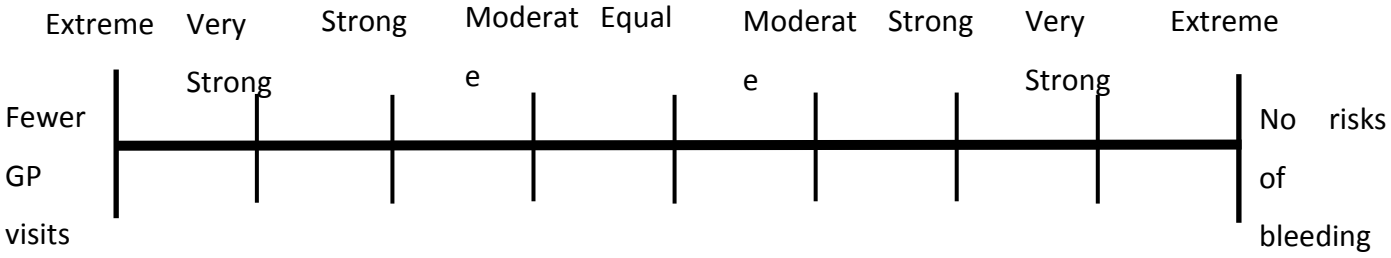
30. What is your preference: Less visits to the GP or no risk of bleeding from a treatment?

Before you decide, please consider these two pieces of information:

A tonsillectomy can mean you make one less visit to the GP in the first 6 months after surgery.

There is a small chance of bleeding following a tonsillectomy. Bleeding is not fatal but would require you to return to hospital for special medications.

Please indicate your answer by placing a cross on the scale below



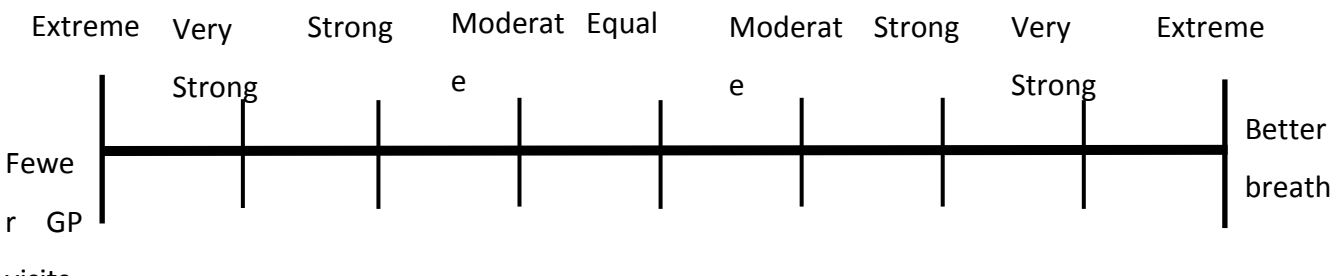
<u>31.</u> What is your preference: Fewer visits to the GP or having better breath?

Before you decide, please consider these two pieces of information:

A tonsillectomy can mean you make one less visit to the GP in the first 6 months after surgery.

Some people who had bad breath from the recurring sore throats reported better breath following tonsillectomy.

Please indicate your answer by placing a cross on the scale below



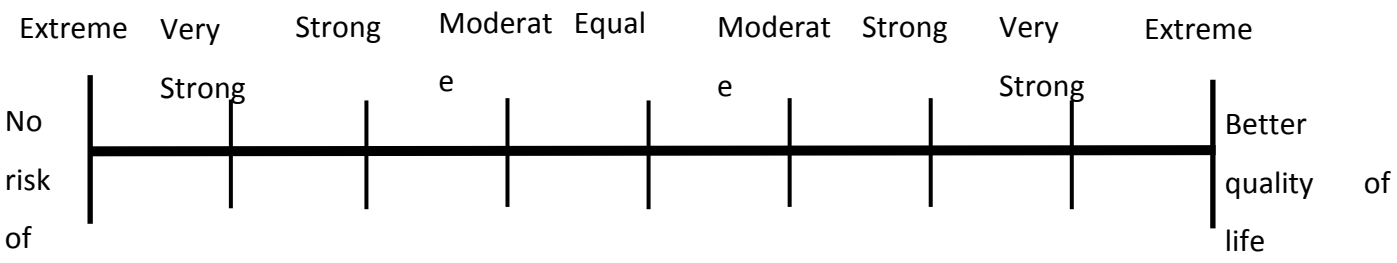
32. What is your preference: Better quality of life or No risk of bleeding?

Before you decide, please consider these two pieces of information:

After a tonsillectomy, some patients report a large improvement in their quality of life

There is a small chance of bleeding following a tonsillectomy. Bleeding is not fatal but would require you to return to hospital for special medications.

Please indicate your answer by placing a cross on the scale below



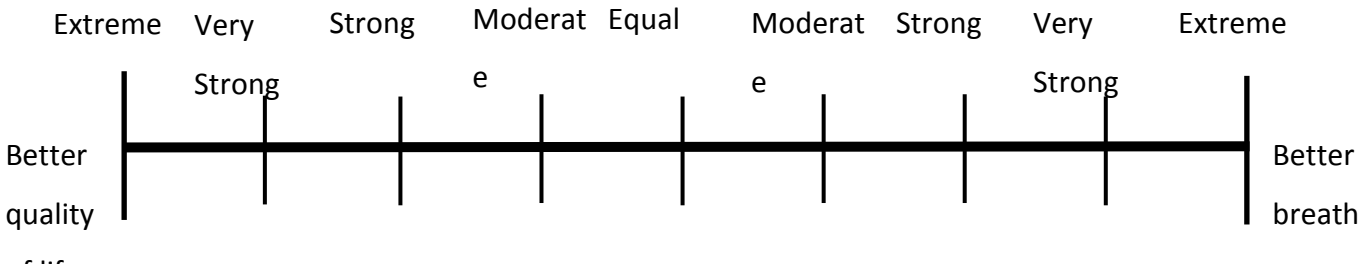
33. What is your preference: Better quality of life or better breath?

Before you decide, please consider these two pieces of information:

After a tonsillectomy, some patients report a large improvement in their quality of life (medium quality research studies)

Some people who had bad breath from the recurring sore throats reported better breath following tonsillectomy.

Please indicate your answer by placing a cross on the scale below



Appendix S – Doctor post consultation questionnaire

Patient Unique Identifier _____ Doctor Unique Identifier _____

Doctor satisfaction and certainty following decisions

Both sets of questions relate to the patient you have just seen and helped reach a decision about which treatment they should undertake to treat their recurrent tonsillitis. The first set of questions are about your certainty during the decision making process. The second set of questions relate to your satisfaction of the decision making process. Please tick only one box per question

1. Which treatment option did you and the patient decide upon?

☐ Tonsillectomy ☐ Watch and Wait

Thinking about the decision you have just made with your patient please consider each of the following statements and tick a box that you most agree with in each row

	Strongly disagree	Disagree	Neither agree or disagree	Agree	Strongly agree
2. The decision was hard to make	☐	☐	☐	☐	☐
3. I was unsure what treatment would really be best for this patient	☐	☐	☐	☐	☐
4. When making the decision, I felt I did not know enough about the treatment alternatives, although the information is available in the literature	☐	☐	☐	☐	☐

5. I had trouble making the decision because important information is either unknown or not readily available in the literature	☐	☐	☐	☐	☐
	Strongly disagree	Disagree	Neither agree or disagree	Agree	Strongly agree
6. When I made a decision it was hard to decide if the benefits of the available treatments were more important than the risks or vice versa	☐	☐	☐	☐	☐
7. It was easy to identify all the considerations that affect the decision	☐	☐	☐	☐	☐
8. I fully understand the patient's views regarding the important issues in making this decision	☐	☐	☐	☐	☐
9. I believe that the patient fully understands the risks and benefits of the treatment we chose	☐	☐	☐	☐	☐

10. I believe that the patient will adhere to the treatment chosen	☐	☐	☐	☐	☐
11. I am satisfied with the decision that was made	☐	☐	☐	☐	☐
12. I am satisfied that the process used to make the decision was as good as can be	☐	☐	☐	☐	☐

The following questions relate to how you reached the above decision. Please tick a box that you most agree with in each row.

	Strongly disagree	**Disagree**	**Neither agree of disagree**	**Agree**	**Strongly agree**
13. I am satisfied that I am adequately informed about the issues important to my decision.	☐	☐	☐	☐	☐
14. The decision I made was the best decision possible for me personally.	☐	☐	☐	☐	☐
15. I am satisfied that my decision was consistent with my personal values.	☐	☐	☐	☐	☐
16. I expect to successfully carry out (or continue to	☐	☐	☐	☐	☐

carry out) the decision I made.					
17. I am satisfied that this was my decision to make	☐	☐	☐	☐	☐
18. I am satisfied with decision	☐	☐	☐	☐	☐

Thank you completing the questionnaire. Please now give the questionnaire to your researcher.

Appendix T AHP methodology and Ranking results

An Eigenvector grid of outcome comparisons was created for each participant. Each row represented an outcome. The matrix was sequentially squared until the normalised row total stabilised. The normalised row total allowed the participants preference for that particular outcome to be quantified in relation to the other four outcomes. All five row totals always summed to one. Participants' ranks for outcomes were created based on their normalised row totals. Therefore, all five outcomes could be ranked in order of priority for the participant from one to five, with five being the most important and one the least. Ranked treatment priorities were described individually for patients and surgeons.

Weights were calculated reflect the likelihood and effect size of each outcome within each treatment option. Weights were created by making pairwise comparisons between tonsillectomy and conservative therapy, for each of the five above outcomes individually. For example tonsillectomy and conservative therapy were compared to each other with regards to the likelihood and effect size of bleeding after each treatment: High level evidence shows that the risk of bleeding following tonsillectomy is between 4-10%, whilst there is 0% risk of bleeding following conservative therapy. Therefore the comparison would 'extremely' favour conservative therapy in this instance. This was repeated for the other four outcomes and the resulting verbal responses converted to the Saaty scale (as described above). Five comparison matrixes were created, with each row representing a treatment option (tonsillectomy or conservative therapy). The square root of the row product was normalised to create a priority vector, or an outcome weight.

The association between Ranked Outcome Priorities (ROP) and treatment chosen

To assess if higher ranking of an outcome changed the treatment chosen I undertook ordinal logistic analyses between patients' ROP for each of the five outcomes (reducing GP visits, reducing days of sore throat, reducing halitosis, improving quality of life, reducing chance of bleeding) and treatment chosen (tonsillectomy or watchful waiting). ROPs that were significantly related to the treatment choice in univariable analyses were sequentially added to a multivariable ordinal logistic regression, allowing for clustering of patients within ENT

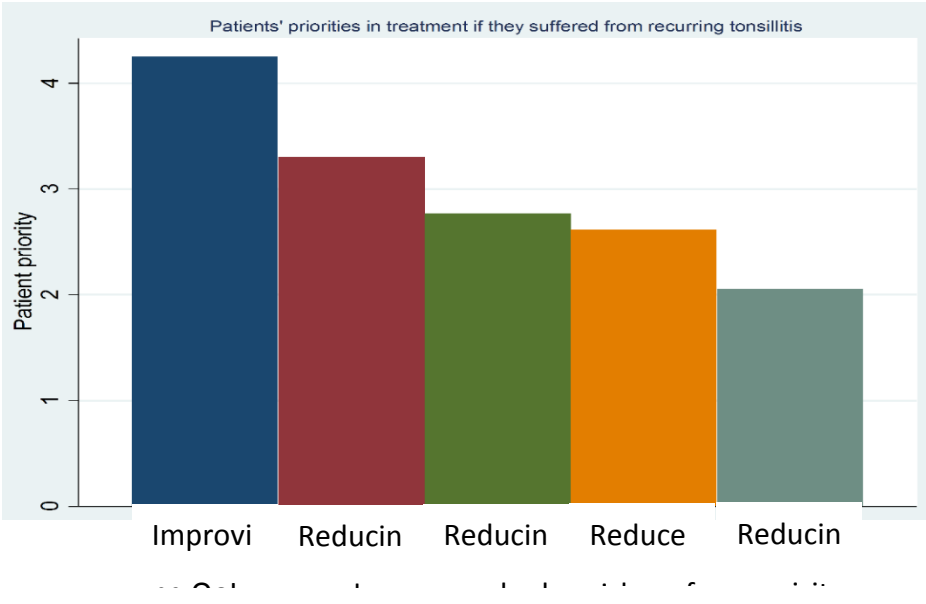
surgeons. Each ROP was added sequentially and retained if it reached statistical significance ($p<0.05$). A multivariable model of patients' rankings was created to control for patient age, sex, ethnicity and originating hospital.

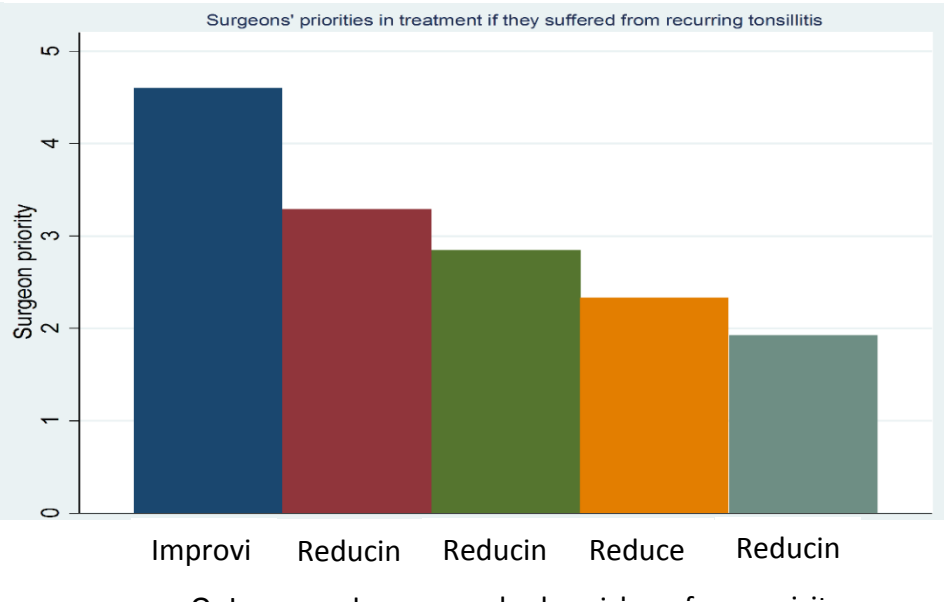
The association between Ranked Outcome Priorities and treatment chosen

Patients' and ENT surgeons' ranking of each outcome was calculated using standard analytical hierarchy process described above. As two groups, ENT surgeons and patients, both ranked potential outcomes identically in terms of their priorities: Patients and ENT surgeons ranked quality of life as the most important outcome in making their decision, with reducing days of sore throat number 2, reducing bad breath number 3, reducing the risk of bleeding number 4 and reducing visits to the GP as the least important. These ranking were the same in the group of patients who selected tonsillectomy as treatment as well as those who selected conservative therapy (higher score shows higher rank and greater preference).

Ordinal Logistic Regression showed that patients' priority ranking of outcomes was not significantly associated with actual treatment choice. Patients choosing tonsillectomy tended to rank improving quality of life highly, however, this result was not statistically significant. Logistic analyses of ENT surgeons' outcome rankings showed that if an ENT surgeon placed a high rank on reducing chance of bleeding that consultation was less likely to end in tonsillectomy ($p=0.03$).

A multivariable model of patients' rankings confirmed that no outcome ranking score was associated with treatment choice, even after accounting for patient characteristics. A multivariable mode of ENT surgeons' rankings showed that after accounting for grade of surgeon, type of originating hospital, surgeon's age and sex there was still a significant association between ENT surgeons who ranked reducing the risk of bleeding high and their patients choosing conservative therapy ($p<0.001$).


Patients' priorities in treatment if they suffered from recurring tonsillitis
Patient priority
0
1
2
3
4
Improvi
Reducin
Reducin
Reduce
Reducin


Surgeons' priorities in treatment if they suffered from recurring tonsillitis
Surgeon priority
0
1
2
3
4
5
Improvi
Reducin
Reducin
Reduce
Reducin

Outcome	Number of patients choosing tonsillectomy with this rank	Number of patients with this rank overall	Ordinal logistic regression OR (95% CI)	P	Number of consultations resulting in tonsillectomy if surgeon ranked this value	Number of consultations where surgeon had this ranking	Firth nominal logistic OR (95% CI)	P2
Reducing day of sore throat								
1 (least important)	14	18	1.75 (0.46-6.72)	0.36	19	24	1	0.29
2	14	18	1.75 (0.46-6.72)		5	5	0.47(0.14-1.66)	
3	35	41	2.92(0.92-9.23)		36	49	1.47(0.07-29.87)	
4	32	38	2.67(0.84-8.47)		48	54	0.36(0.13-1.01)	
5 (Most important)	20	30	1		16	22	0.34(0.1-1.15)	
Reducing visits to the GP								
1 (least important)	37	47	1	0.17	37	45	1	0.69

2	47	55	1.59(0.57-4.42)		64	83	0.75(0.30-1.84)	
3	25	33	0.84(0.29-2.43)		19	21	1.77(0.39-8.02)	
4	4	8	0.27(0.06-1.28)		2	2	1.13(0.05-25.83)	
5 (Most important)	2	2	-		2	3	0.38(0.04-3.27)	
Improving QoL								
1 (least important)	0	1	-	0.16	0	0	-	0.18
2	1	3	0.08(0.01--0.99)		1	1	1	
3	19	25	0.52(0.17-1.62)		2	4	0.33(0.01-12.82)	
4	34	45	0.51(0.20-1.31)		37	50	0.93(0.04-24.13)	
5 (Most important)	61	71	1		15	84	1.82(0.07-46.68)	
Reducing Bad breath								
1 (least important)	32	39	1	0.5	17	27	1	0.16
2	24	31	0.75(0.23-2.43)		25	31	2.36(0.74-7.45)	

3	16	23	0.5(0.15-1.67)		43	51	3.07(1.06-8.87)	
4	25	28	1.82(0.43-7.77)		24	29	2.67(0.80-8.88)	
5 (Most important)	18	24	0.66(0.19-2.25)		15	16	6.20(0.98-39.13)	
Reducing risk of bleeding								
1 (least important)	32	40	1	0.89	51	58	1	0.03
2	29	38	0.81(0.27-2.36)		29	34	0.78(0.24-2.57)	
3	20	23	1.67(0.39-7.03)		24	29	0.65(0.20-2.16)	
4	20	26	0.83(0.25-2.76)		13	19	0.30(0.09-1.01)	
5 (Most important)	14	18	0.88(0.23-3.39)		7	14	0.15(0.04-0.52)	

Surgeon variables	No of consultations resulting in tonsillectomy	Number of consultations altogether	Adjusted OR	P

Rank of "reduce risk of bleeding"				
1(Least important)	51	58	1	<0.001
2	29	34	0.49(0.05-4.9)	
3	24	29	0.78(0.17-3.48)	
4	13	19	0.39(0.07-2.20)	
5(Most important)	7	14	0.05(0.01-0.18)	
Age				
20-29	7	8	1	0.81
30-39	47	58	0.81(0.08-8.56)	
40-49	45	61	0.87(0.05-15.22)	
50-59	35	41	1.03(0.06-16.90)	
60-69	8	10	1.12(0.07-17.99)	
Gender				

Male	101	126	1	0.63
Female	41	52	1.71(0.19-15.33)	
Originating Hospital				
District General Hospital	101	144	3.52(1.09-11.37)	0.04
University hospital	43	46	1	
Grade				
Consultant	72	93	1	0.01
Registrar	36	44	5.54(1.51-20.40)	
Staff Grade	11	15	1.13(0.19-6.64)	
Associate Specialist	24	23	4.86(0.91-25.93)	
Core trainee	2	3	1.22(0.07-22.37)	

Appendix U - Sore throat codes

readcode	readterm	medcode
14B7.00	History of recurrent tonsillitis	95893
1C9..00	Sore throat symptom	5755
1C9..11	Throat soreness	404
1C92.00	Has a sore throat	5553
1C93.00	Persistent sore throat	12489
1C9Z.00	Sore throat symptom NOS	15287
1CB3.00	Throat pain	386
1CB3.11	Pain in throat	7366
2DB..11	O/E - tonsils enlarged	18539
2DB2.00	O/E - tonsils hyperaemic	22131
2DB3.00	O/E - tonsils mod. enlarged	6498
2DB4.00	O/E - tonsils grossly enlarged	10291

2DB5.00	O/E - tonsils - quinsy present	24596
2DB6.00	O/E - follicular tonsillitis	7266
2DB7.00	O/E - exudate on tonsils	25176
2DC1.00	O/E - pharynx hyperaemic	24664
2DC2.00	O/E - granular pharyngitis	24788
2DC3.00	Inflamed throat	14931
2DE7.00	O/E - throat haemorrhage	71829
4JF4000	Throat swab culture positive	27014
7531100	Drainage of peritonsillar abscess	6596
A34..00	Streptococcal sore throat and scarlatina	54777
A340.00	Streptococcal sore throat	1765
A340200	Streptococcal pharyngitis	4902
A340300	Streptococcal tonsillitis	8496
A340z00	Streptococcal sore throat NOS	16217
A34z.00	Streptococcal sore throat with scarlatina NOS	16184

A383000	Fusobacterial necrotising tonsillitis	58538
A772.00	Viral pharyngoconjunctivitis	27324
A912300	Primary tonsil syphilis	48291
A913400	Secondary syphilis of tonsils	37158
A986.00	Gonococcal pharynx infection	50882
AA12.00	Vincent's pharyngitis	31536
AA1z.12	Vincent's tonsillitis	16954
AA25.11	Rhinopharyngitis mutilans	53708
AB63100	Tonsillar aspergillosis	16543
H00..00	Acute nasopharyngitis	3260
H02..00	Acute pharyngitis	893
H02..11	Sore throat NOS	6014
H02..12	Viral sore throat NOS	6466
H02..13	Throat infection - pharyngitis	310
H020.00	Acute gangrenous pharyngitis	36219

H021.00	Acute phlegmonous pharyngitis	24708
H022.00	Acute ulcerative pharyngitis	21486
H023.00	Acute bacterial pharyngitis	17899
H023000	Acute pneumococcal pharyngitis	92428
H023100	Acute staphylococcal pharyngitis	29589
H023z00	Acute bacterial pharyngitis NOS	53395
H024.00	Acute viral pharyngitis	4868
H025.00	Allergic pharyngitis	6274
H02z.00	Acute pharyngitis NOS	407
H03..00	Acute tonsillitis	138
H03..11	Throat infection - tonsillitis	11499
H03..12	Tonsillitis	2125
H030.00	Acute erythematous tonsillitis	12010
H031.00	Acute follicular tonsillitis	4061
H032.00	Acute ulcerative tonsillitis	8452

H033.00	Acute catarrhal tonsillitis	37409
H034.00	Acute gangrenous tonsillitis	59986
H035.00	Acute bacterial tonsillitis	10156
H035000	Acute pneumococcal tonsillitis	58188
H035100	Acute staphylococcal tonsillitis	64973
H035z00	Acute bacterial tonsillitis NOS	15970
H036.00	Acute viral tonsillitis	9357
H037.00	Recurrent acute tonsillitis	1747
H03z.00	Acute tonsillitis NOS	20104
H12..00	Chronic pharyngitis and nasopharyngitis	10083
H121.00	Chronic pharyngitis	4324
H121.11	Sore throat - chronic	16814
H121000	Simple chronic pharyngitis	47426
H121100	Atrophic pharyngitis	38879
H121200	Granular pharyngitis	21562

H121300	Hypertrophic pharyngitis	15794
H121400	Pharyngitis keratosa	56361
H121500	Pharyngitis sicca	30569
H121600	Chronic follicular pharyngitis	47269
H121z00	Chronic pharyngitis NOS	14926
H122.00	Chronic nasopharyngitis	12667
H12z.00	Chronic pharyngitis and nasopharyngitis NOS	54657
H14..00	Chronic tonsil and adenoid disease	21000
H14..12	Tonsil disease - chronic	16864
H140.00	Chronic tonsillitis	1667
H141.00	Tonsil and/or adenoid hypertrophy	3549
H141.12	Enlargement of tonsil or adenoid	18238
H141000	Hypertrophy of tonsils and adenoids	24164
H141100	Hypertrophy of tonsils alone	2158
H141z00	Hypertrophy of tonsils and adenoids NOS	29502

H143.00	Chronic adenotonsillitis	9328
H14y.00	Other chronic diseases of tonsils and adenoids	54475
H14y500	Caseous tonsillitis	36462
H14y600	Lingular tonsillitis	35249
H15..00	Peritonsillar abscess - quinsy	3605
H1y2.00	Other pharyngeal disease NEC	10355
H1y2100	Pharynx or nasopharynx cellulitis	25156
H1y2200	Parapharyngeal abscess	12231
H1y2300	Retropharyngeal abscess	27279
H1y2400	Pharynx or nasopharynx oedema	19948
H1y2600	Pharynx or nasopharynx abscess	14710
H271100	Influenza with pharyngitis	29617
Hyu0100	[X]Acute pharyngitis due to other specified organisms	93964
Hyu0200	[X]Acute tonsillitis due to other specified organisms	73118

J083600	Uvulitis	8480
R041.00	[D]Throat pain	15039
R041.11	[D]Throat discomfort	21060
14B6.00	History of quinsy	96011
2DB5.11	O/E - quinsy present	6971
2DC1.11	O/E - fauces injected	22396
4JH5000	Mouth swab culture positive	44211
7531111	Drainage of quinsy	7956
H15..11	Quinsy	911
Hyu2500	[X]Other chronic diseases of tonsils and adenoids	3430

Appendix V – Table of Interrelatedness of Concepts

Concept name	Involvement	Decisional support	Information exchange	Role preference	Empowerment	Uncertainty	Treatment preference

	Full Definition as per original text from articles	**1.**Patient as person; **2.**Explore all biopsychosocial causes; **3.**Patient involvement in decisions; **4.**Responsibility for non medical aspects of care; **5.**Giving information to patients, **6.**Form therapeutic alliance;	**1.**Assess patient characteristics for success; **2.** Information delivery; **3.**Value elicitation; **4.**Support decision maker according to conflict; **5.**Evaluate quality of decision	**1.** Physician and patient roles, **2.** Goals needs and expectations, **3.**Treatment options; **4.**Explain cause of problems; **5.**Explore patients ideas **6.**Explore patients concerns;	**1.**Beliefs in the efficacy and benefits of self-care; **2.**Choice of role decision making **3.** Communicating role preference	**1.** The need to respect individuals decision making abilities and to recognise their capacity to make those decisions; **2.** The need for workers to surrender their need for control and link in with support networks in a co-operative	**1.**Information gathering about gains and losses from change in status quo **2.** Information gathering about gains and losses about alternatives **3.** Time to evaluate above	**1.**Outcome information gathering **2.**Weighing likelihood of outcomes versus preference for outcomes

		7. Share power and responsibility; **8.**Equality of doctor-patient relationship		**7.**Explore patients expectations; **8.**Discuss effect of problem on patient's life; **9.**Explain treatment; **10.**Discuss side effects and risks; **11.**Discuss decision on treatment		and collaborative manner; **3.** Identification of the power imbalances in relationship; 4.awareness that the user may reject help offered; **5.** The need for workers to secure and use the resources that will promote or foster a sense		

						of control and promote individual ability; **6.** The adoption of a person valuing approach		
Involvement/centredness	**1.**Patient as person; **2.**Explore all biopsychosocial causes; **3.**Patient involvement in decisions; **4.**Responsibilit		1-1 2-1 3-3 5-4 5-2 6-4 7-4	1-1 1-2 2-4 2-8 3-2 3-5 3-6	1-1 3-2 4-2 5-3 6-2	1-1 7-2 7-3 7-5 8-6	2-1 5-1 5-2	2-1 5-2

	y for non medical aspects of care; **5.**Giving information to patients, **6.**Form therapeutic alliance; **7.** Share power and responsibility; **8.**Equality of doctor-patient relationship		8-4	3-7 5-9 5-10				

Decisional support	**1.**Assess patient characteristics for success; **2.** Information delivery; **3.**Value elicitation; **4.**Support decision maker according to conflict; **5.**Evaluate quality of decision	[redacted]	1-2 1-5 1-6 1-7 2-2 2-4 2-8 2-9 2-10 3-3 3-5 3-6 3-7 5-11	1-1 3-3 4-2	1-3 3-6 4-5	3-1 3-2	3-1 4-2

Information exchange	**1.** Physician and patient roles, **2.** Goals needs and expectations, **3.** Treatment options; **4.** Explain cause of problems; **5.** Explore patients ideas **6.** Explore patients concerns;	1 & 7	10	Physician and patient roles	1-2 1-3	1-3 2-6 5-6 6-6 7-6	5-1 5-2 6-1 6-2 7-1 7-2 8-1 8-2	5-1 6-1 7-1 8-1 10-1 10-2

	7. Explore patients expectations; **8.** Discuss effect of problem on patient's life; **9.** Explain treatment; **10.**Discuss side effects and risks; **11.** Discuss decision on treatment					

Role preference	**1.**Beliefs in the efficacy and benefits of self-care; **2.** Choice of role decision making **3.** Communicating role preference	[redacted]	1-1 2-1	NIL	NIL
Empowerment	**1.** The need to respect individuals decision making abilities and to recognise their	[redacted]		NIL	NIL

	capacity to make those decisions; **2.** The need for workers to surrender their need for control and link in with support networks in a co-operative and collaborative manner; **3.** Identification of the power imbalances in relationship;	[redacted]		

	4. Awareness that the user may reject help offered; **5.** The need for workers to secure and use the resources that will promote or foster a sense of control and promote individual ability; **6.** The adoption of a person valuing approach	[redacted]		

		[redacted]		
Decisional uncertainty	**1.**Information gathering about gains and losses from change in status quo **2.** Information gathering about gains and losses about alternatives	[redacted]	[redacted]	1-1 1-2 2-1 2-2

	3. Time to evaluate above	[redacted]	

This table takes the numbered definitions of SDM constructs in the row (blue) and shows how they have overlap with numbered definitions of SDM constructs in the column (red). For example, code 2-1 was entered for row of involvement and column decision support. It meant the 2nd numbered definition of the row construct (involvement) – that is “explore all biopsychosocial causes” - was considered to overlap with the 1st numbered definition of the column construct (decisional support) – that is “assess patient characteristics for success”.

Appendix W – Gold standard psychometric properties

Psychometric property	Definition/test	Criteria for acceptability
Item reduction	Identification of items for possible elimination owing to weak psychometric performance; assessed on the basis of (1) unrotated principal component factor analysis to determine whether all items are measuring a single factor; and (2) item analyses for all items	Principal component factor analysis All items should load on the first unrotated factor >0.30 Item analyses (applied to all items): Missing data <5% No item redundancy (inter-item correlations ≤ 0.75) Item–total correlations >0.25 Maximum endorsement frequencies ≤ 80% (i.e. the proportion of respondents who endorse each response category), including floor/ceiling effects <80% (i.e. response categories with high endorsement rates at the bottom/top ends of the scale, respectively)

		Aggregate adjacent endorsement frequencies ≥ 10 (391)
Acceptability	The quality of data; assessed by completeness of data and score distributions	Missing data for summary scores <5% Even distribution of endorsement frequencies across response categories Floor/ceiling effects for summary scores <10%
Reliability		
Internal consistency	The extent to which items comprising a scale measure the same construct (e.g. homogeneity of the scale); assessed by Cronbach's alpha(392)	Cronbach's alphas for summary scores ≥ 0.70 Inter-item or Item–total correlations ≥ 0.20(393)
Test–retest reliability	The stability of a measuring instrument; assessed by administering the instrument to respondents on two different occasions and examining the	Test–retest reliability correlations for summary scores ≥ 0.70 (394)

	correlation between test and retest scores	
Inter-rater reliability	Agreement between independent raters/observers; assessed by ICCs	ICC ≥ 0.70(394)
Parallel (alternative) forms reliability	Agreement between two or more parallel/alternative forms or different versions of the same measure (e.g. form A/B, short/long form) that indicates that they can be used interchangeably; assessed on the basis of correlations between parallel/alternative forms of a measure	High correlation between parallel/alternative forms of the measure (e.g. between long and short form)
Validity		
Content validity	The extent to which the content of a scale is representative of the conceptual domain it is intended to	Qualitative evidence from pre-testing with patients, expert opinion and literature review that items in the

	cover; assessed qualitatively during the questionnaire development stage through pre-testing with patients, expert opinion and literature review	scale are representative of the construct being measured
Criterion-related validity **Concurrent validity**	Evidence that the scale predicts a gold-standard criterion that is measured at the same time; assessed on the basis of correlations between the scale and the criterion measure	High correlation between the scale and the criterion measure
Criterion-related validity **Predictive validity**	Evidence that the scale predicts a gold-standard criterion that is measured in the future; assessed on the basis of correlations between the scale and the criterion measure	High correlation between the scale and the criterion measure
Construct validity **Within-scale analyses**	Evidence that a single entity (construct) is being measured and that items can be combined to form a summary score; assessed on the basis	Internal consistency (Cronbach's alpha) ≥ 0.70 Moderate to high correlations between scale scores

	of evidence of good internal consistency and correlations between scale scores (which purport to measure related aspects of the construct)	
Construct validity **Convergent validity**	Evidence that the scale is correlated with other measures of the same or similar constructs; assessed on the basis of correlations between the measure and other similar measures	Correlations are expected to vary according to the degree of similarity between the constructs that are being measured by each instrument. Specific hypotheses are formulated and predictions tested on the basis of correlations
Construct validity **Discriminant validity**	Evidence that the scale is not correlated with measures of different constructs; assessed on the basis of correlations with measures of different constructs	Low correlations between the instrument and measures of different constructs

Construct validity **Known groups differences**	The ability of a scale to differentiate known groups; assessed by comparing scores for subgroups who are expected to differ on the construct being measured	Significant differences between known groups or difference of expected magnitude
Responsiveness	The ability of a scale to detect clinically important change over time; assessed by comparing scores before and after an intervention of known efficacy (on the basis of various methods including *t*-tests(395) effect sizes(309) standardised response means(396) or responsiveness statistics(397)	Significant differences between known groups or difference of expected magnitude

Appendix X – Psychometric properties of SDM instruments

Year	Name of Instrument	Item reduction	Acceptability	Content validity	Reliability			Construct validity				Responsiveness
					Internal consistency Cronbach's α	Test-retest	Inter-rater reliability	Convergent validity	Discriminant validity	Known group differences	Within scale analyses	Experimental intervention
2010	Dyadic OPTION	Not reported	Not reported	Qual evidence Expert opinion	No data	No data	Pearson's correlation coefficient between doctor and observer =0.58 (p<0.001)	Not reported	Patient or doctor gender did not affect scores, values not provided	No data	No data	No data
2001	FPICS	No data	No data	Expert opinion	0.93	Retested over 10 weeks Pearson's R =0.85	Not relevant	Correlates with: Patient Satisfaction Questionnaire (r=0.67) Patient Communication Style Scale (r=0.38)	Does not have differing results according to patient age, gender, and level of education (t-test p>025))	Did not pick up suspected difference when doctors are female	EFA: 1 factor None of the items loadings was less than Cronbach's alpha	No data

								General Adherence Scale (r=0.32) Length of time with doctor (r=0.17) General Health Perceptions Scale (r=0.19)			0.72 on the primary factor	
2010	SDM-Q9	<20% missing data for any item Item-total correlations>0.69	No data	Qual evidence Expert opinion Lit review	0.94	No data	Not relevant	Correlates with Control Preference scale r=0.48 (398)	Items do not discriminate age, sex, education level, health problem in consultation or topic of consultation	Can discriminate between increasing levels of SDM in simulated encounters, p<0.001 (399)	EFA: 1-factorial structure All items loading on this factor: Cronbach's alpha=0.90	No data
1990	PPICS	No data	No data	Qual evidence	0.73	No data	Not relevant	Ware satisfaction with decision scale (r= 0.26, p<0.05)	Scores do not change with age Women report higher scores (PCC 0.39, p=0.03)	Can discriminate between increasing levels of SDM in simulated encounters, p<0.001	EFA: 3 factors Factor 1 Cronbachs alpha of 5 items >0.50, explaining 11% of the variaince Factor 2 Cronbachs alpha	No data

										(399)	of 4 items >0.53 explaining 25% of the variance Factor 3 Cronbachs alpha of 4 items >0.37 explaining 10% of the variance	
2006	HCEQ	Item-total correlation all 0.3-0.67	No data	No data	0.83	Retested at 16 days ICC=0.70	Not relevant	No data	No data	No data	EFA: 3 factors explaining 69% of the variance. Cronbach's alpha>0.6 for all items belonging to a factor and <0.51 if not belonging to that factor	No data

1995	DCS	No data	<1%missingness	Expert opinion Lit review	0.92	Retest done at 2 weeks Pearson's R=0.81	Not relevant	Perceived risk questionnaire (r=-0.47) Knowledge tests (r=-0.16) Perceived Utility of Genetic testing scale (r=-0.3)(400) Perceived information levels (aOR=0.46) (401)	Scores do not change with age	Difference was found between those who chose a treatment and those who delayed Also difference between those choosing to undergo screening and those who didn't (402)	EFA 4 factors explaining 71% variance) (402)	Decision aids reduce DCS scores by mean of 7.26 (Stacey)

Table 42 Psychometric properties of SDM instrument

Appendix Y – Tonsillectomy Outcomes Evidence and Patient Ranking

List of factors	Study type	Effect size	Effect small/med/large	Study strength	Average patient rank	Rank
Days of a sore throat	Meta-analysis of RCT	3.61/6months	1. large	1. Strong	3.6	3
Number of episodes of sore throat	Meta-analysis of RCT	10.64/6months	1. large	1. Strong	2.2	1
Visits to the GP	RCT	0.9/6months	1. large	1. Strong	4.6	4
QoL	Case series	10% improvement in physical component sf 36	2. medium	3. Weak/Moderate	1.8	1
Snoring reduction	Cohort	46% reduction in odds of being a severe/habitual snorer if you had a tonsillectomy	1. large	3. Weak/Moderate	5.2	8
Voice change	Case control	17% of male children had hypernasalance preop, and 9% had it post tonsillectomy	2. medium	3. Weak	10..3	11
Halitosis reduction	Case series	80% had clearance of halitosis at 2 months	1. large	3. Weak/Moderate	4	5

Societal cost	Case series	12.3 year break even time	2. large	4. weak	8.2	10
Taste disturbance	Case series	10%/6months	2. medium	3. Weak/Moderate	7	6
Haemorrhage risk	Cohort	4.9% incidence adults (all indication tonsillectomy)	1. large	2. Moderate/Strong	7	6
Immunological profile	Case series	Mild reduction in IgM	3. small	3. Weak/Moderate	8.4	9

Table 43 Tonsillectomy outcomes evidence and patient rankings

Appendix Z – Work in progress

Do latent surgeon treatment preferences influence treatment decisions in adults with recurring tonsillitis?

Introduction

There is strong evidence of regional variations in the tonsillectomy rates, with little evidence on whether patients' preferences for tonsillectomy vary by region. Some authors have suggested that regional variations in tonsillectomy are the results of regional variation in surgeon preferences for tonsillectomy. However, there is very little evidence of whether and how surgeons influence treatment decisions at the level of the consultation. In this study we aimed to evaluate the role of patient and surgeon treatment preferences in the choice between tonsillectomy and conservative therapy in adults with recurring tonsillitis.

Methods

Using the Preferences in Adults with Recurring Tonsillitis Tool (PARTT), we undertook a multicentre (n=14) observational study of consecutive adults attending ENT clinics with recurring tonsillitis. Surgeons (n=54) at participating hospitals were asked to also complete PARTT before they saw any patients, from the perspective of a typical adult with recurring tonsillitis – surgeon's proxy preference. Patients (n=160) who consented to participate completed the PARTT following their consultation, as well information regarding the treatment they had chosen.

We undertook cluster analyses to assess the impact of surgeons' proxy preference against the patients'. Multivariable logistic models were created to assess the impact of preference cluster on treatment actually chosen.

Results

We analysed data from 160 consultations between 160 patients and 54 surgeons. PARTT responses fell into three clusters: aligning tonsillectomy (n=66 patients, n=16 surgeons), aligning conservative therapy n=22 patients, n=8 surgeons) and undecided (n=72 patients, n=30 surgeons).Multivariable logistic analysis showed that treatment chosen at the end of the consultation was not related to patients' preference cluster (p=0.48). However, surgeons'

preference cluster was associated with treatment chosen, even after controlling for markers of disease severity and patient preference cluster (Adjusted OR 3.88, 95%CI 1.01-14.97).

Discussion

This study adds weight to the argument that surgeon factors may be more consequential in treatments chosen than patients' preferences. In the modern setting where conditions frequently have several therapies, all in equipoise, it is important that treatments chosen reflect patient's personal values. Future work should investigate methods that allow consultation outcomes to reflect patients' preferences more closely. Using the PARTT routinely, would help to make patients values more explicit whilst making surgeons more aware of their implicit preferences that may be unintentionally swaying treatment decisions.