IHS Statistics Team

Statistical Analysis Plan for the NATTINA Trial

The **NA**tional **T**rial of **T**onsillectomy **IN A**dults: a clinical and cost effectiveness study

SAP Version number: 3.0

ISRCTN Number: 55284102

Trial Statistician:

Signature:

Date: 26/02/21

Reviewed and Approved by:

Professor Signature:

Date: 02 March 2021

Professor J **Signature**:

Date: 01 March 2021

This SAP and preceding versions will be stored in the Statistics Trial Master File.

Document history

Version	Date	Changes made	Justification for change
Version 1.0	5.1.17	Completed version	
Version1.1	Nov 2017 March 2019	Sensitivity analysis maybe removing weeks soon after surgery Instrumental variables	Sore throat days in weeks after surgery could affect differences in total number over 2 years data statistical co-applicant recommendation
Version 1.2	Dec 2019	Final draft	Discussion with statistical co-applicant
Version 1.3	Jan 2020	Updated Final draft	New version of protocol (v7.0)
Version 1.4	Jan 2020	Updated Final draft	Discussion with statistical co-applicant and senior statistician
Version 2.0	June 2020	Population groups for analysis clarified and Per protocol group added Section 8 primary analysis clarified, sensitivity analysis specified and per treatment and per protocol analyses specified	Final version to ensure analysis of final trial data is clear
Version 2.0_clar	8th Aug 2020	Clarification of the agreed primary analysis (8th August 2020)	Additional clarification of the primary analysis. Agreed by TMG via email
Version 3.0	26 Feb 2021	Track changes removed	To upload the SAP onto the funder Management Information System website

1. INTRODUCTION

This statistical analysis plan (SAP) provides guidelines for the analysis and presentation for the analysis of the NATTINA trial and is based on the protocol version 7.0 (11th September 2019). This plan, along with all other versions and associated documents relating to the analysis of this trial, will be stored in the 'Statistical Documentation' of the Trial Master File. The final version of the SAP will be in place before the final comparative analysis of the trial data is undertaken.

A selection of dummy tables and figures relating to this plan can be found in the Appendix.

1.1 Trial Summary

Short title:	NATTINA
Chief Investigator:	Professor Janet Wilson
Sponsor:	The Newcastle upon Tyne Hospitals NHS Foundation Trust
Funder:	Health Technology Assessment, Clinical Evaluation and Trials
Study design:	A multi-centre randomised controlled surgical trial with internal pilot.
Study Intervention:	1:1 randomisation of immediate tonsillectomy versus conservative management (i.e. deferred surgery)
Primary outcome:	Total number of sore throat days over the 24 months following randomisation.
Number of study sites:	28 (was initially 9)
Study population/size:	510 patients will be recruited in total.
Study duration:	57 months

1.2 Study design

A multi-centre, randomised, controlled surgical trial. Patients are randomised into two parallel streams on a 1:1 ratio stratified by centre and baseline severity. The patient's severity category is determined by the total TOI 14 score from the Baseline Questionnaire Package, as follows: Mild = 0 to 35, Moderate = 36 to 48 and Severe = 49 to 70

Participants are adult patients with acute tonsillitis who have been referred to otolaryngology outpatient clinics for recurrent sore throat.

1.3 Trial Objectives

The purpose of the study is to establish the clinical and cost effectiveness of tonsillectomy compared with conservative management for adult tonsillitis which, through observation and statistical modelling of outcomes, will evaluate the impact of alternative sore throat patient pathways and develop future research.

Primary Objective:

To compare the effectiveness (as number of sore throat days) and efficiency of tonsillectomy versus nonsurgical management for recurrent acute tonsillitis over the 24 months following randomisation.

Secondary Objectives:

i) Clinical Effectiveness:

- To compare other metrics of sore throat severity including responses on the Tonsil Outcome Inventory 14 and weekly STAR response data for any sore throat episodes experienced.
- To compare quality-of-life as recorded by SF-12 longitudinally during study follow up.
- To report the number of adverse events, visits to the GP/walk-in clinic/A&E, prescriptions issued and additional interventions required for sore throats and related events through weekly STAR response data, and supported by data linkage to primary care patient records.
- To adjust the estimate of effectiveness in light of other baseline covariates including severity of tonsillitis.
- To evaluate the impact of alternative sore throat patient pathways by observation and statistical modelling of outcomes.
- To assess to what extent trial participants are representative of the total population of sore throat patients referred to ENT clinics.

ii) Economic Evaluation:

- To compare QALYs using the AUC method based upon SF-6D scores derived from the SF-12 responses measured at baseline, throughout the study and during any episodes of sore throat experienced.
- To compare the cost-effectiveness measured in terms of the incremental cost per sore throat day avoided from the perspective of the NHS and patients over 24 months
- To compare the cost-utility based on incremental cost per QALY gained from the perspective of the NHS and participants over 24 months
- To compare the cost-benefits based on the perspective of the NHS and participants' willingness to pay to avoid a sore throat day using the NATTINA contingent valuation questionnaire 'Value of Avoiding a Sore Throat Day' administered at baseline

iii) Qualitative Process Evaluation: To document the views, experiences and acceptability of patients and clinicians regarding tonsillectomy and conservative management, and how patient experience may shape future research required

iv) Future Research: To propose further research questions using newfound cost-benefit information to develop algorithms that guide and assess management of health services.

1.4 Sample size

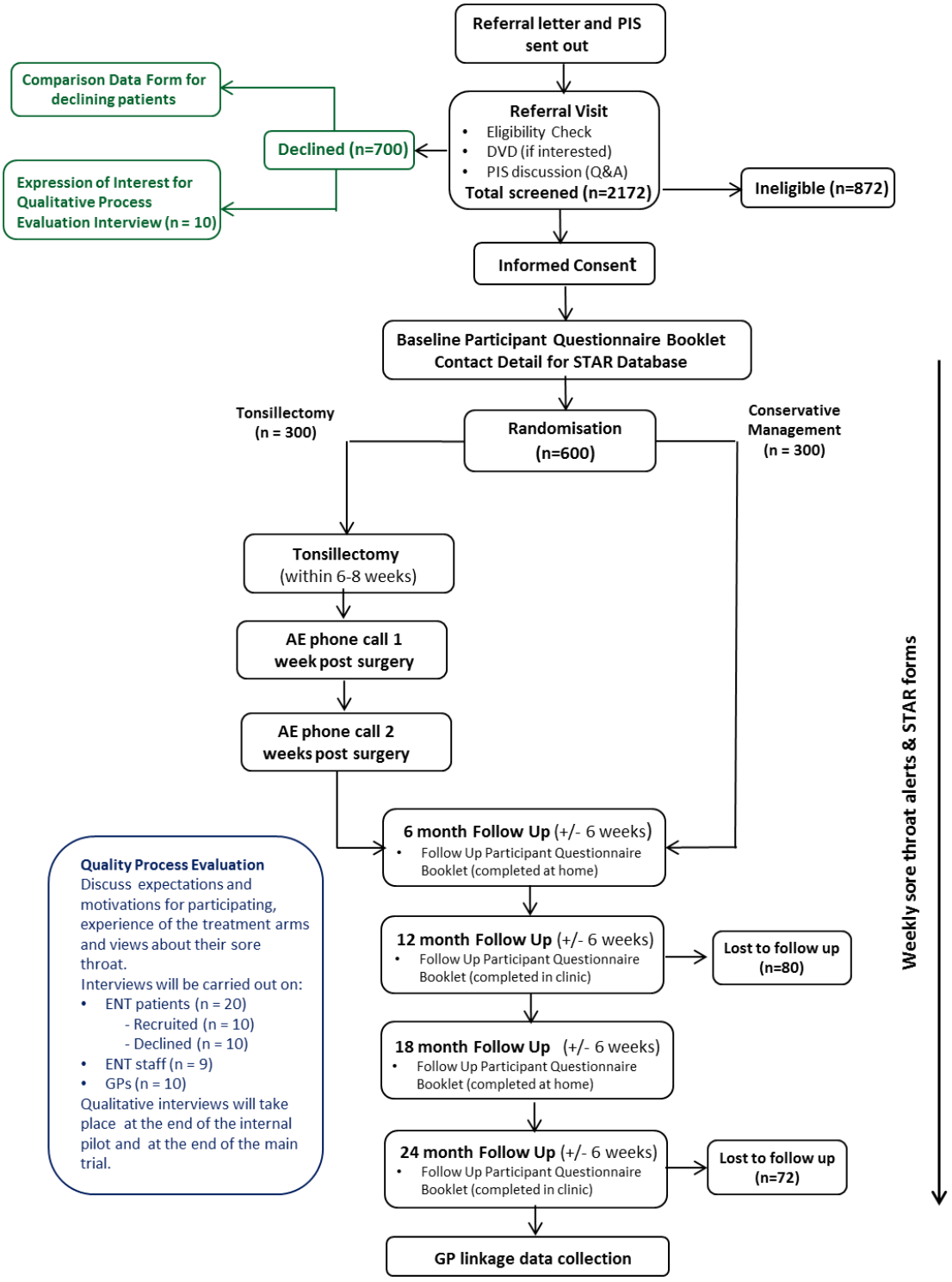
A total of 510 participants (255 in each arm) will be recruited throughout the study. Note that the internal pilot phase of the trial was cancelled after consultation with HTA.

Extract From protocol

The total number recruited will be 510 including 72 in the internal pilot. By recruiting 510 patients we are allowing for a total loss to follow up of 25% over 24 months. 382 patients in

total two groups of 191 patients (providing complete data at two years) gives 90% power to detect an effect size of 0.33 (corresponding mean intergroup difference of 3.6 days of sore throat based on a pooled estimated standard deviation of 10.8 days) assuming a type 1 error rate of 5%. The sample size calculations take account of the anticipated losses as well as predicted switch rates. We anticipate that our loss to follow-up rate should be less than the stated 25%, as we shall intensively follow-up trial participants in both arms.

1.5 Trial Diagram/Flowchart



2. TIMING AND REPORTING INTERIM AND FINAL ANALYSES

2.1 Analysis timelines

2.1.1 Declaration of formal stopping rules to be implemented within trial

None specified.

2.1.2 Final analysis timeline based on planned FU and/ or events

Final analysis is planned for summer 2020; the qualitative analysis is planned for January 2017.

There are no formal interim analyses of the primary outcome measure planned except for snapshots reported to DMC. DMC/TSC meetings are held annually, but may be held more frequently if requested.

HTA reports are submitted biannually. Data may be provided to HTA if requested in the interim.

The final analysis will be carried out once all recruitment and data collection including follow up has been completed, data cleaned and database locked. Data sets will be merged according to unique patient identifier allocated at randomisation.

Data snapshots are downloaded from MACRO into statistical software package STATA v15

3. RECRUITMENT AND RANDOMISATION

3.1 Recruitment

- Trial status: (*recruiting or closed to recruitment*)
- Grant awarded:
- Ethics awarded:
- Number of sites:
- Date first site open:
- Date first participant randomised:
- Date of snapshot:
- Total number of participants randomised before snapshot:
- Date last participant randomised (before snapshot):

3.2 Recruitment Summary

Recruitment will be reported in a consort flow diagram. Figures showing recruitment over time and cumulative number of patients randomised by month will be provided.

A table showing expected and actual recruitment will be provided. A figure will be produced based on this table.

A table with summarised site information will be provided including number of sites open and number of sites that have randomised at least 1 patient at the time of the particular report. A further table showing the cumulative number of sites open by month will be provided.

A table showing the distribution of participants by site and randomised trial arm will be provided.

3.3 Randomisation

A blocked allocation (permuted random blocks of variable length) system is used to allocate subjects to the 2 intervention groups; tonsillectomy versus conservative management, in a 1:1 ratio stratified by centre and severity. Randomisation is administered centrally via the NCTU using a secure web-based system, accessed by the PI or delegated individual. The patient's severity category is determined by the total TOI 14 score from the Baseline Questionnaire Package, as follows: Mild = 0 to 35, Moderate = 36 to 48, Severe = 49 to 70.

3.4 Confirming balance across trial arms by stratification

A table to confirm the balanced number of patients in each arm by strata for the ITT population will be provided to show balance by site & baseline severity.

3.5 Blinding

No blinding in this trial

3.6 Ineligible Patients

Despite treatment withdrawal, patients will continue to be followed in the study. The primary statistical analysis will be carried out on an intention to treat basis, retaining patients in their randomised treatment groups and including protocol violator and ineligible patients. Ineligible patients are classed as those randomised patients who are found to subsequently not adhere to the eligibility criteria of the trial. The number of ineligible patients and reasons for ineligibility will be reported and a sensitivity analysis may be conducted and reported if the number of ineligible patients is excessive.

Table showing reasons for ineligibility will be provided.

Protocol violators will be reported as part of treatment compliance

3.7 Recruitment of non-randomised (declining) patients

Patients who are eligible but decline to participate will be invited to provide anonymised baseline comparison data for the NATTINA database (age, gender, an estimate of number of sore throat days over the prior 6 months and a TOI 14 questionnaire). This will allow an analysis of the comparability of our trial participants to the total pool of those referred, at each of the sites. Declining patients will also be invited to participate in a qualitative interview with a researcher. Further details of the analysis involving non-randomised patients can be seen in Section 5 dealing with the study populations.

4. DATA QUALITY

4.1 Forms Returned

Data are collected using case report forms (CRF). Completion rates for each CRF will be reported. CRF's are completed and collated in the following order:

i) Baseline participant questionnaire and randomisation at referral/baseline visit
ii) Weekly STAR response - Sore throat returns (weekly text sent from baseline to month 24)
iii) Additional STAR questionnaire for any week with sore throat returns greater than zero
iv) Surgery CRF (within 6-8 weeks of baseline/randomisation) including information phone calls;
    - 1 week after surgery (post- operative phone call)
    - 2 weeks after surgery (post- operative phone call)
v) Follow up participant questionnaire by post (6 months +/- 6weeks)
vi) Follow up participant questionnaire at visit (12 months +/- 6weeks)
vii) Follow up participant questionnaire by post (18 months +/- 6weeks)
viii) Follow up participant questionnaire at visit (24 months +/- 6weeks)
ix) Serious Adverse Event forms
x) Cross over form for participants switching treatment
xi) Withdrawal form

A table will be presented that gives the expected numbers, actual numbers and response rate for the trial CRF's. This will also include the primary outcome measure collected via the weekly STAR response texts (ii) and resultant STAR questionnaires when sore throat days>0. These are sent by each participant on a weekly basis following randomisation. Each non-zero weekly STAR response (i.e. participants reporting between 1 and 7 sore throat days) triggers a reminder for the participant to complete a STAR questionnaire with more details of these sore throat days provided.

4.2 Assessment of visit & postal completion of questionnaires compliance

Where feasible, visit (& postal return) windows of +/- 6 weeks should ensure sufficient time is offered to facilitate scheduling appointments; non-attendance for study visits will prompt follow-up by telephone. Participants may also be contacted via telephone by the research nurse at the participating site to remind or encourage them to return questionnaires or weekly alerts. Source data verification will be performed by the Trial Manager at each participating site.

4.3 Charts showing time in days between visits

Charts showing time in days between randomisation and each follow up visit with reference lines added to show the compliance window will be provided. This will be of the form of a histogram with compliance reference lines added at +/- 6 weeks

5. STUDY POPULATION

5.1 Defining population for analysis

The primary statistical analyses will be carried out on an intention to treat basis, retaining patients in their randomised treatment groups and including protocol violator and ineligible patients. However patients may switch over from conservative to surgical management. In the NATTINA design, patients are asked to commit to "deferred surgery". We anticipate that a number of patients will take up this opportunity to switch to surgery. The implication of such crossover, which typifies surgical trials, is that the intention to treat analysis will produce a very conservative estimate of the effect of tonsillectomy. We will therefore also undertake a per treatment (as treated) analysis with repeated measures corresponding to two periods of follow up for those patients who crossover from medical management to tonsillectomy.

- Intention to treat group (ITT) – all ineligible and protocol violator participants will be included in the analysis on an intention to treat basis with participants kept in their randomised treatment group.
- Per treatment group – all randomised participants who start treatment included in the analysis according to the treatment they receive.
- Per protocol as randomised- limited to patients who were randomised to surgery and received within the 8 week window and those patients randomised to the conservative arm who did not crossover.
- Non-randomised group - those eligible to be included in the trial but declining to take part

5.2 Baseline Patient Characteristics

The population is intention to treat (ITT)

Demographic and baseline characteristics and trial stratification factors (site and baseline severity) at randomisation will be compared across treatment groups descriptively. Descriptive statistics will be tabulated by treatment group and overall.

A summary of the health and well-being section of the study questionnaires will be included here. This includes self-rated health, work and activity limitation and health service utilisation.

Demographic characteristics for comparison will be presented.

A second table reporting ITT population and non-randomised participants will also be provided.

Trial factors for comparison will include stratification factors (site and baseline severity) and time in days from baseline visit to randomisation.

Self-rated health at baseline (from SF12) will be tabulated by arm and overall

6. TREATMENT RECEIVED

6.1 Tonsillectomy

In NATTINA, referral of patients to ENT by GPs for consideration of tonsillectomy follows the current standard care pathway according to NICE guidelines. Consenting participants who are eligible for elective tonsillectomy are randomly allocated to one of two arms; elective surgery (identical to that in standard care) and conservative management.

In the event a participant is randomised to undergo a tonsillectomy but surgery is delayed due to severe tonsillitis or other complications, the participant should remain in the trial and continue to follow the surgery pathway.

Details of the surgery, including date and any complications experienced up to 30 days after the tonsillectomy will be documented in the CRF. These will be presented in the analysis reports.

1 week and 2 weeks after surgery

The research nurse will contact the participant twice after their tonsillectomy to check on their recovery and ask if they have experienced any adverse events immediately after, or during recovery from, a tonsillectomy. Only participants who had surgery will be contacted. AEs and SAEs will be recorded. Compliance for the two phone calls are presented in section 4 and adverse events in section 7.

Clinic visits 2 and 3 – 12 and 24 month follow up (+/- 6 weeks)

All participants are reviewed in the outpatient clinic at 24 months. This is the final review. Compliance is reported in section 2.

The analysis set is the per treatment set.
Charts showing time in days between randomisation and tonsillectomy with reference lines added to show the compliance window will be provided. This will be of the form of a histogram with compliance reference lines added at +/- 6 weeks

Note for the compliance for the Tonsillectomy, we note that surgery is preferable within 6 weeks of randomisation, but the maximum time elapsed to still comply with the protocol is 8 weeks. An additional reference line at +8 weeks will be added to the tonsillectomy compliance chart to take account of this.

The analyses will be repeated as for the ITT group who had surgery at any time.

6.2 Withdrawal and Crossover of Participants

Participants have the right to withdraw from the study at any time for any reason, and without giving a reason. The investigator also has the right to withdraw participants from the study intervention if he/she judges this to the in the patient's best interests. Should a participant decide to withdraw from the study, all efforts will be made to report the reason for withdrawal as thoroughly as possible.

There are 2 options for participants in the Immediate Tonsillectomy group:

1. Cross over to conservative management before receiving the intervention (surgery) - continue with follow up visits and data collection under conservative management pathway.
2. Withdraw completely before or after surgery – no further follow up visits or data collection will occur.

There are 2 options for participants in the Conservative Management group:

1. Withdraw completely from study - no further data will be collected.

2. Cross over to surgery – continue with the follow up visits and data collection as scheduled. Those who still meet the SIGN guidelines for tonsillectomy will be put forward for surgery.

All data collected up until withdrawal will be retained for NATTINA research purposes.

The numbers of withdrawals will be presented along with reasons and time from randomisation.

The protocol states an expected dropout rate of 25%. The sample size calculation is based on this figure so we will comment on the dropout rate and assess the implications of the actual observed rate with any actions required if expected rate is exceeded.

The numbers of participants who crossed over from their randomised arm will be presented along with reasons and time from randomisation when this occurred.

Patients who do cross over will be censored at the time of crossover for Intention to treat group and included in the per treatment group from the date of crossing over.

Non-compliance (including crossover) may be addressed using an 'as treated' approach or complier average causal effect (CACE) approach, since intention to treat analysis under non-compliance is biased when the intervention effect is large (Dumville 2006). Alternative analysis can provide less biased estimates (Chenglin 2014).

Statistical methods for withdrawal of patients, based on statistical censoring, will be considered. The crucial aspect to these proposed analyses is collating information on date and reason for withdrawal or crossover. This was addressed in a previous TMG and data collection modified accordingly.

6.3 Concomitant Medications

Information will be gathered and reported on whether any over the counter or prescription medication have been used via both the STAR questionnaires and at the 24 month final follow up visit.

7. SAFETY ANALYSIS

Adverse events (AE) are graded according to section 16 of the trial protocol.

- Adverse events must be related to the study intervention.
- Adverse events will be collected and recorded at the 2 post-operative phone calls at 1 and 2 weeks after surgery.
- Any serious adverse events will be recorded throughout the duration of the trial until the 24 month follow up and once they are resolved.
- Serious adverse events exclude any pre-planned hospitalisations (e.g. elective surgery) not associated with clinical deterioration.
- Serious adverse events exclude routine treatment or monitoring of the studied indication, not associated with any deterioration in condition.
- Serious adverse events exclude elective or scheduled treatment for pre-existing conditions that did not worsen during the study.

All adverse events related to the study intervention are reported

The research nurse will contact participants twice after their tonsillectomy (1 week and 2 weeks after surgery). The purpose of the contact is to check on the patient's recovery and to ask if they have experienced any adverse events immediately after, or during their recovery from the tonsillectomy. Only participants that have had surgery will be contacted and reported.

Severity of AEs will be graded on a three-point scale (mild, moderate, severe). Relation (causality) and seriousness of the AE to the treatment will be reported. All AE's, SAE's and adverse reactions will be reported by the number of events (n) in unique patients (N). They will be reported in the following sequence;
Definitely related to treatment – Severe, moderate then mild
Probably related to treatment – Severe, moderate then mild
Possibly related to treatment – Severe, moderate then mild

All SAEs during study participation will be reported to the DMC as line listings including relationship of the SAE to study procedures and expected (according to the protocol) or unexpected nature of the SAE. The number of treatment related serious adverse events (SAE), including treatment related deaths, are reported divided by their relationship as 'definitely', 'probably' and 'possibly' related to treatment.

Safety data will not be subject to statistical analysis.

8. EFFICACY ANALYSIS – PRIMARY OUTCOME

8.1 Measuring STAR Response

Patient population is the ITT group

Number of sore throat days

All participants submit weekly feedback on the number of sore throat days experienced over the previous 7 days. This is in the weekly STAR response texts (range 0 to 7)

Primary outcome measure

The primary outcome measure is the total number of sore throat days collected by weekly STAR response text (range from 0 to 7 each week), experienced over the 24 months of follow-up. This outcome data will be analysed at the end of the study.

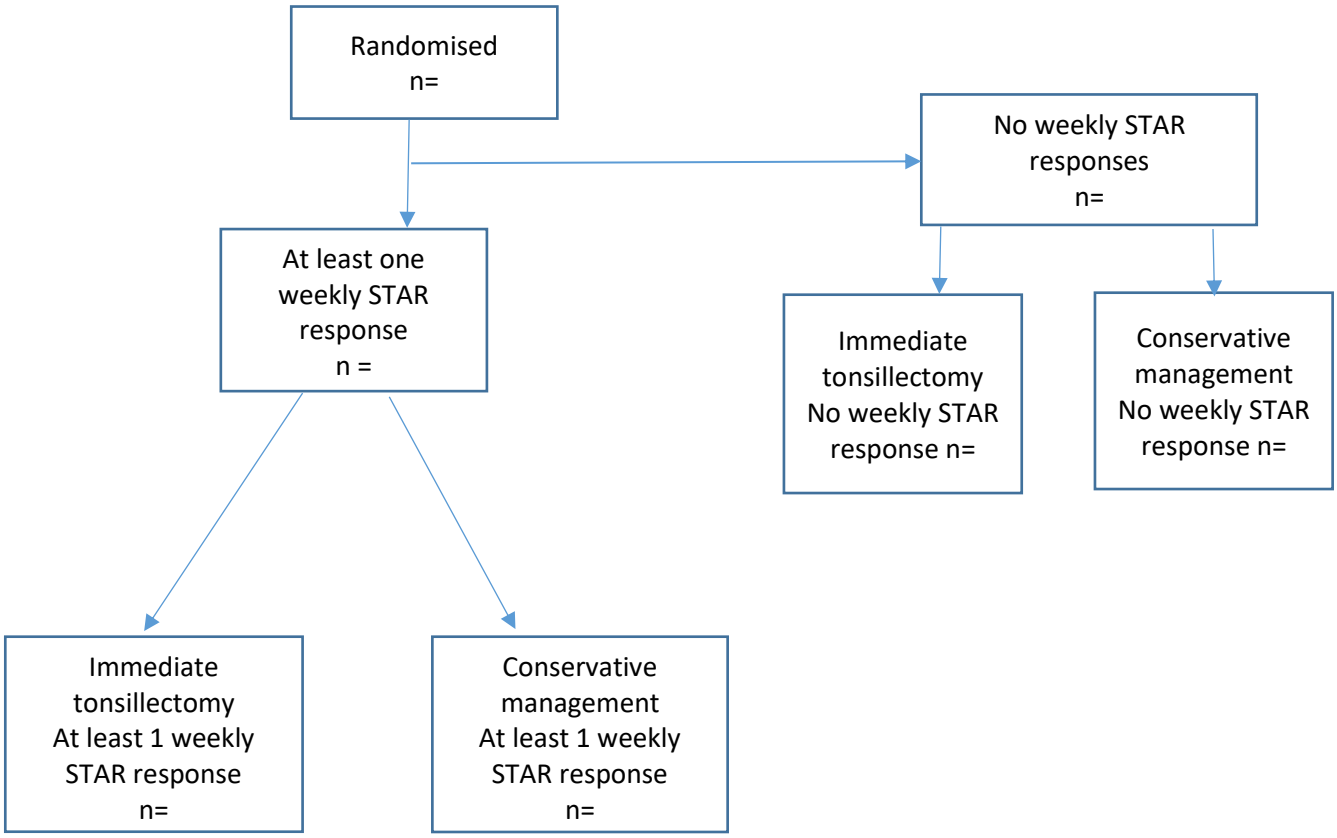
The STAR database records response (referred to as weekly STAR response) as a single number. A participant who has suffered from a sore throat in the past week (i.e sore throat days > 0) is asked to complete a supplementary STAR Questionnaire. The STAR questionnaires provide additional data for health economics and other secondary outcomes (see section 9).

8.2 Descriptive Analysis of primary outcome measure

Median (IQR) and overall range will be presented for the average time on trial (weeks) by arm for participants who have made at least one weekly STAR response (N).

8.2.1 Retention flow diagram

Flowchart 1



Flow chart 1 shows the number of randomised patients and how they have responded to the weekly STAR response procedure. We will show whether the response is balanced across treatment arms.
The TMG has implemented a robust system of weekly monitoring of sore throat STAR responses. The database provider (Inteleme) provides the trial managers with weekly updates on the total number of responses, the STAR questionnaires requested and any missed response. Participants that consecutively (2 or more times) do not supply a weekly STAR response are flagged to sites and are contacted by site staff. These highlight the importance of the weekly STAR response to the participants

Retention is defined as the time on the trial calculated as the time from date of randomisation to date of last weekly response. The average time on the trial for the participants who have sent at least one weekly STAR response will be presented in box plots: this will allow us to assess if retention appears to be balanced by treatment arms.

8.2.2 Description of weekly STAR response counts over time

A chart showing observed counts of weekly STAR responses from when the participants enter the trial (week 0) up to the end of the 2 year follow up period (104 weeks) will be provided. Comparative charts separate for each treatment arm will also be presented to highlight any differences in return rates that may exist between arms.

8.2.3 Number of sore throat days reported over the 2 year follow up period
The cumulative total number of sore throat days reported on a patient level will be provided separately for each treatment arm (for surgery split up into pre surgery, surgery + 4 weeks, and rest of 2 year follow up period, dummy table provided in appendix). A time series chart (Randomisation to 104 weeks) will also be provided with the point of tonsillectomy highlighted for those receiving surgery to allow comparison of rate before and after the operation, both in the initial four postoperative weeks (related to surgery) and until the end of follow-up. The point of surgery and weeks after that where we would expect sore throat days to be reported will be indicated on the graph (for example a histogram showing time of surgery after randomisation).

8.2.4 Differential Loss to follow up of primary outcome measure
We will report the differential loss to follow up in the following way.

Of N patients randomised, there are n patients to date who have not completed any weekly STAR response, stating whether this is balanced across arms.

The table will include study ID, arm randomised to, date randomised, time on trial in weeks and whether they have withdrawn or not.

8.2.5 Missing primary outcome data
Weekly STAR response can be categorised at each time point as i) complete, where patients have completed all of the expected weekly STAR responses within that time frame, ii) partial, where some but not all expected weekly STAR responses have been received and iii) no returns, where patients have not returned any of the expected weekly STAR responses within that time frame.

A table will be provided that shows the expected number of responses, observed complete, observed partial and missing (no weekly STAR response returns) for each of 6 months, 12 months, 18 months and 24 months when other trial questionnaires are due to be completed. The table will also include median and interquartile ranges of the numbers of weekly STAR response at each time point. Times are measured from randomisation.

A bar chart will be presented that shows the expected number of weekly STAR response returns split by full, partial and missing as described above, split by treatment arm. This will demonstrate whether response rates appear balanced across randomised treatment groups.

Attrition can introduce bias if the characteristics of the patients lost to follow-up differ (Dumville 2006). Internal validity depends on the balance of characteristics across randomised arms in patients who are retained. A table will be provided that details the characteristics of patients across randomised treatment arms according to complete, partial or no weekly STAR response at the time points.
An assessment of whether characteristics of patients responding and not responding appear balanced by randomised treatment arm will also be provided.

Addressing Missing data

Patients with missing observations will be examined to determine both its extent and whether it is missing at random or is informative (i.e. crossover, lost to follow up, etc.). If data is missing to a sufficient extent and considered missing at random, the use of appropriate multiple imputation techniques may be considered as secondary analysis.

8.3 Primary analysis

This primary analysis will be undertaken on an intention to treat basis.

The data will be analysed using negative binomial regression. This is appropriate as count data is usually analysed using poisson regression, but in situations where the data is over dispersed (variance much bigger than mean) negative binomial regression is more robust. Count data often have an exposure variable indicating length of time observed. Exposure variables can be applied to the negative binomial regression model developed to account for differences in the proportion of the 104 weekly returns completed by each participant.

The univariate analysis of the overall treatment effect will be the unadjusted negative binomial regression analysis without taking account of any stratification variables. Response variable is the total number of sore throat days over 24 months from randomisation to last follow up.

The primary hypothesis to be tested is H_0: number of sore throat days in 24 months is same for both treatment arms. A two-sided significance level of $p<0.05$ will be used throughout.

A multivariable (adjusted) analysis will compare the NATTINA arms while adjusting for the stratification variables used at the point of randomisation in the trial - recruiting centre (as a random effect) and baseline severity (as a fixed effect).

Both unadjusted and adjusted analyses will be undertaken. The adjusted analysis will be the key primary analysis.

8.4 Sensitivity analysis

Further multivariable analyses will consider other important baseline factors in the negative binomial regression model including gender, age, ethnicity, education level, employment status, site and baseline levels. Nonlinear continuous covariates will be transformed where appropriate using simple first degree polynomial transformations. Exposure variable will be incorporated into the negative binomial regression model developed.

8.5 Additional analyses of the Primary Outcome

8.5.1 Addressing Crossover (Per treated analysis)

Patients may switch over from conservative to surgical management. The implication of such crossover, which typifies surgical trials, is that the intention to treat analysis will produce a conservative estimate of the effect of tonsillectomy. We may also undertake an analysis on the per treatment group with repeated measures corresponding to two periods of follow up for those patients who crossover from medical management to tonsillectomy. The length of these follow up periods will be as an exposure variable in the negative binomial regression.

Simple adjustment methods such as censoring switchers at the point of switch, may be prone to selection bias should the prognosis be different in those who switch treatments (for example they may have a worse sore throat). Various statistical methods are available to adjust estimates in the presence of treatment switching, but each makes important assumptions and is subject to limitations, but may be explored. These include i) the inverse probability of censoring weighting (ICPW) method which, as with per protocol analysis, assumes that outcome data collected after a switch is irrelevant and follow up data are censored at the time of switching. However this method must include all variables that predict both treatment switching and outcome which may be difficult to achieve; ii)

Marginal Structural models (MSM) compares treatment histories - treated at start (identified from the treatment arm), never treated (identified from the weighted placebo arm, censored at treatment), treated for x months. The model is estimated by weighting the data to estimate outcomes under each potential treatment history, so the time of switching can be included; Rank preserving structural nested failure time models (RPSFTMs) estimate the crossover-corrected treatment effect where many patients in a placebo arm switch to treatment arm, but can be complicated by censoring. We plan to consider the methods outlined and will make a decision as to which is the most appropriate during the final analysis. The selection of the method will be dependent on the scale and pattern of switching observed in the trial.

8.5.2 Instrumental Variables

ITT principle keeps the participants in the arm they were randomised to regardless of the treatment they actually receive. Consequently ITT analysis does not capture the full efficacy of surgery trials where there is non-compliance with treatment arms originally randomised to.[1] Instrumental variables rely on building a model that predicts the treatment actually received, accounting for unexpected behaviour between variables related to the treatment received. We will investigate instrumental variable analysis in the presence of crossover at lost to follow up levels predicted at design stage if the number crossing over from their allocated treatment is excessive.

1. **Complier average causal effect (CACE)**

An approach that better estimates the effect of the treatment than either per protocol, where participants that comply with allocated treatment are analysed, or per treatment, where treatment actually received is analysed, is the complier average causal effect (CACE)[2]. CACE analysis allows unbiased assessment of treatment effect, after grouping the intervention arm into compliers and non-compliers[3]. Complier average causal effect (CACE) can be reduced for estimation to **CACE= ITT/proportion of compliers in treatment arm[4]** Where CACE is defined as mean of observed outcomes for participants who comply with allocated treatment.

2. **Interactions with treatment arm allocation**

In RCT the instruments include the randomised arm and interactions between arm and baseline covariates[3]. As an exploratory analysis we will model and report parameter estimates of interaction term only of treatment arm and baseline severity.

8.5.3 Addressing 8 week Surgery subgroup (Per protocol analysis)

[1] Sitlani, C.M., Heagerty, P.J., Blood, E.A. and Tosteson, T.D. (2012), Longitudinal structural mixed models for the analysis of surgical trials with noncompliance. Statist. Med., 31: 1738-1760. doi:10.1002/sim.4510

[2] Hewitt, C. E., Torgerson, D. J., & Miles, J. N. (2006). Is there another way to take account of noncompliance in randomized controlled trials?. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*, *175*(4), 347. doi:10.1503/cmaj.051625

[3] Bond, Simon J, White, Ian R, and Sarah Walker, A. Instrumental Variables and Interactions in the Causal Analysis of a Complex Clinical Trial. *Statistics in Medicine.* 26.7 (2007): 1473-496

[4] Dunn, G., Maracy, M. & Tomenson, B. Estimating treatment effects from randomized clinical trials with noncompliance and loss to follow-up: the role of instrumental variable methods. Statistical Methods in Medical Research 2005 14:4, 369-395

Analyses will be restricted to those randomised to and having surgery within the specified 8 weeks following randomisation, compared to those randomised to the conservative arm who did not cross over the surgery.

9. EFFICACY ANALYSIS – SECONDARY OUTCOME MEASURES

Secondary Outcome Measures are:

- Responses on the Tonsil Outcome Inventory 14 (TOI14) and STAR questionnaire data to measure frequency, severity, health and economic impact of any sore throat episodes experienced.
- Longitudinal quality-of-life (as recorded by SF-12 and derived SF 6D) during study follow up using the standardised AUC method
- The number of adverse events, visits to the GP/walk-in clinic/A&E, prescriptions issued and additional interventions required as collected from GP records and other primary care linkage data.
- Incremental cost per sore throat day avoided from the perspective of the NHS and patients over 24 months to measure the cost effectiveness.
- The views and experiences of patients and clinicians regarding tonsillectomy and conservative management and how patient experience may shape any future research required.

9.1 STAR Questionnaires

Sore Throat Alert Return (STAR questionnaire)

A subject who experiences a sore throat is asked to submit a NATTINA STAR questionnaire – comprising:

i. Information on the severity category of sore throat days (mild/moderate/severe)

ii. Report of over-the-counter and prescription medications used

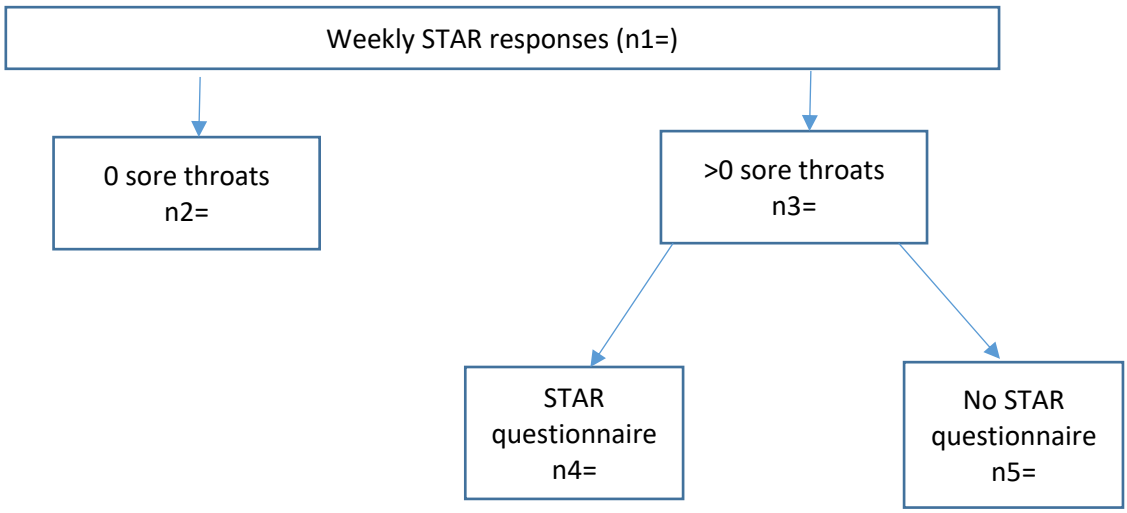
iii. The nature of any professional healthcare advice sought if any (including GP, walk in clinic, pharmacist etc.)

iv. Number of hours unable to undertake usual activities (including time off work and studies)

v. An additional SF-12 relative to the episode

Flowchart 2 shows the questionnaire completion details:

Flowchart 2



In total, N patients have returned a total of n1 weekly STAR responses, n3 of these being non-zero response indicating some level of sore throat. For these n3 non-zero episodes, a STAR questionnaire is also expected (secondary outcome measure) detailing specifics regarding the sore throat incident.
Within the data snapshot, n4 of n3 (x%) STAR questionnaires will have been received and will be reported.

STAR response and STAR questionnaire are mapped by patient ID and date of week.

Responses will be reported descriptively as:
Severity of sore throats
Each time a STAR questionnaire is submitted the patient is asked how bad their sore throat was (mild/moderate/severe) over the 7 day period covered by the form. A table summarising the severity of sore throats will be provided, reported overall and split by arm

Medication
Information on medication used over the 7 day period covered by each STAR questionnaire is collected and will be presented in the health economist report.

Health care advice sought
Information on whether they have sought healthcare advice or attended a healthcare service is also collected via STAR questionnaire returns and will be presented in the health economist report.

Hours of missed work or daily activities
Information collected on amount of time missing from paid work or from doing usual daily activities (hours) over the 2 year follow up period and will be presented in the health economist report

9.2 SF-12 questionnaire

The analysis set is the ITT set. We may look at the other populations if a specific reason emerges during analysis.

Quality of life scores based on the SF-12, will be calculated according to the scoring manual at baseline and 6, 12, 18 and 24 months post randomisation.
Scores will be described (with 95% CI) and graphically presented over time.
Scores will be statistically analysed using longitudinal repeat measure maximum likelihood models developed for longitudinal data. The dependent variable will be the overall quality of life score for an individual patient at a particular occasion (transformed if non-normal). Both variation between patients and variation between responses nested within patients will be modelled as random effects with a normal distribution. Differences between groups and changes over time will be modelled as fixed effects. The analysis will be adjusted for the treatment groups and stratification factors.

Missing SF-12 data will be assessed to decide whether to use complete case analysis or impute missing data. Only if missing at random and not excessive. If missing data is excessive we will describe the situation.

9.3 Tonsillectomy outcome inventory 14 (TOI 14)

The TOI 14 is a validated disease-specific instrument for measuring health-related quality of life and our experience of using the TOI 14 in 3 centres pre and post tonsillectomy equips us to 1) precisely estimate the effect size of tonsillectomy; 2) estimate the spectrum of baseline severity of those referred from primary care for consideration for surgery; 3) account for such variation in the design and analysis of the trial; 4) evaluate the impact of alternative sore throat patient pathways by observation and statistical modelling of outcomes. 'Preop' was removed from the TOI 14 title for the participant questionnaires and Comparison Data Form as they will be used both before and after surgery. The TOI 14 text size and spacing has been marginally modified in the Comparison Data Form so as to ensure it is more user friendly for respondents.

TOI 14 questionnaire data is collected at each for trial participants and also collected for declining patients.

With 14 questions each scoring between 0 and 5 means that the range of scores is 0 to 70 with interpretation (Mild = 0 to 35, Moderate = 36 to 48, Severe = 49 to 70).

Participant population is ITT.

Baseline, 6-month, 12-month, 18-month and 24-month responses to the TOI 14 will be summarised using appropriate descriptive statistics along with 95% CI but will not be statistically tested across treatment groups. We will however test the difference between the means at each time point with confidence intervals.

Similar table showing TOI14 subscales of Throat Discomfort, General Health, Resource impact, Social and psychological will also be provided. Subdomains are throat discomfort (TOIq1+TOIq2+TOIq3+TOIq4), general health (TOIq5+TOIq6), resource impact (TOIq7+TOIq8+TOIq9+TOIq10) and social psychological (TOIq11+TOIq12+TOIq13+TOIq14

Note that after Q14 participants are asked for other symptoms, then description and score given for that. We will check that all who provide another symptom have a score and report along with TOI 14 score. The 'other symptoms' will not be incorporated into the TOI14 scores but will be reported separately.

9.4 18 month Follow up

The participant is asked for information relating to their last hospital admission, hospital appointment, GP/Practice nurse consultations and visits to a walk-in-centre. Information is collected on their method of transport (including distance and cost), the activity they would have been doing if not attending, waiting and travel time and if they were accompanied by a relative or carer.
Descriptive statistics only.

9.5 Health service utilisation

According to Health economics analysis plan. Will not be included as part of the statistical analysis of the trial.

10. APPENDIX: EXAMPLE DUMMY TABLES/ TEMPLATES

Section 3 Recruitment

CONSORT Flow Diagram

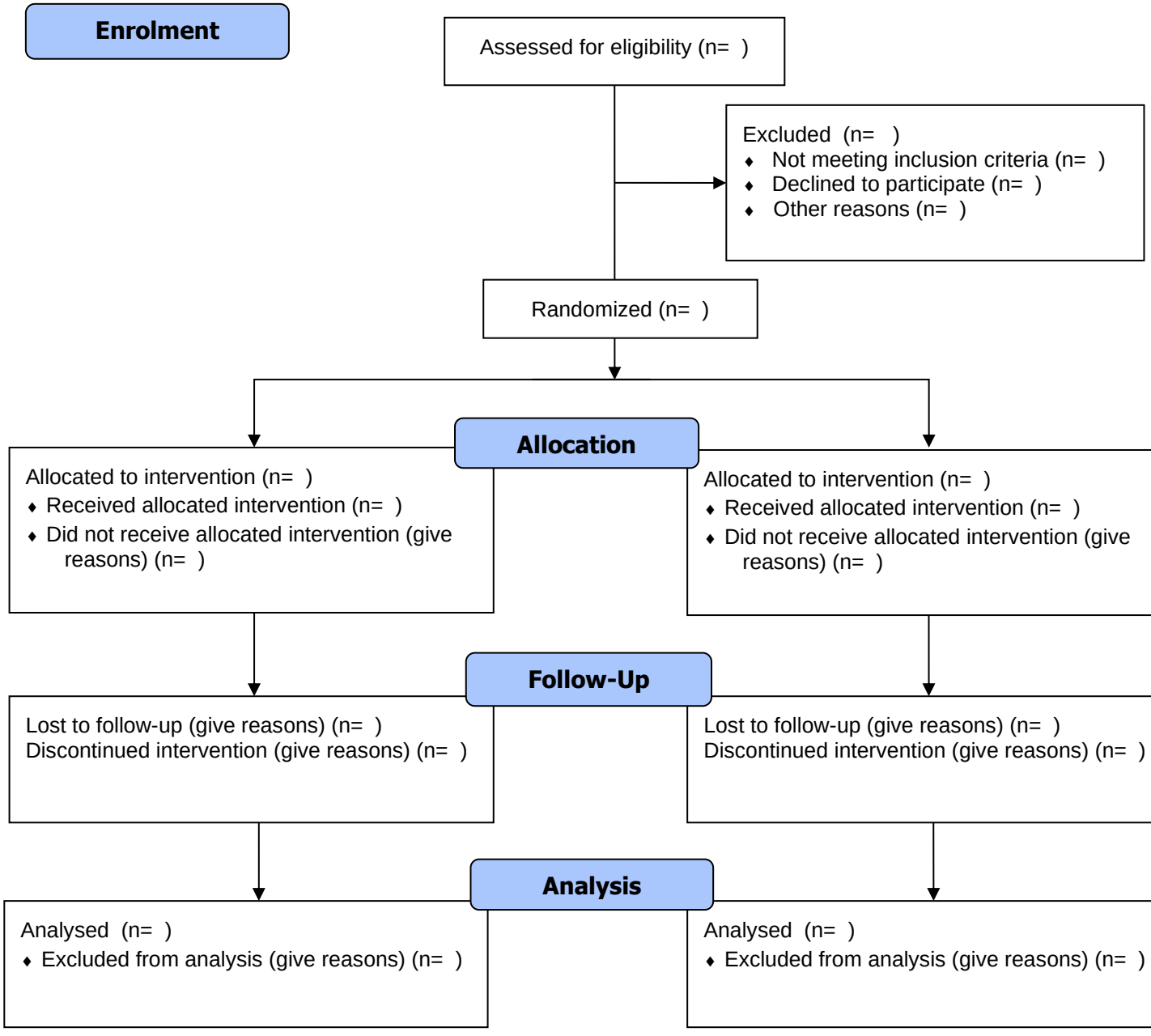


Table 3.1: Site Summary

Number of sites open	n
Number of sites randomised at least 1 patien	n

Table 3.2: Cumulative number of sites open by month

Month	Cumulative number of sites
Date (month 1)	n

Table 3.3 Distribution of participants by site and randomised treatment arm

Site	Randomised treatment arm		Total
	Tonsillectomy	Conservative management	

Table 3.4: Confirm balanced numbers of patients in each arm by strata for ITT population

	Tonsillectomy	Conservative management	Total
Site 1	n (%)	n (%)	n (%)
Site 1	n (%)	n (%)	n (%)
…..	n (%)	n (%)	n (%)

Table 3.5: Distribution of baseline severity (Also confirming whether balanced numbers of patients in each treatment group by stratification factor baseline severity as per the baseline TOI 14 score) for ITT population

baseline severity (by TOI 14 score)	**Immediate Tonsillectomy**	**Conservative management**	Total (N=)
Mild (0-35)	n (x%)	n (x%)	n (x%)
Moderate (36-48)	n (x%)	n (x%)	n (x%)
Severe (49-70)	n (x%)	n (x%)	n (x%)
Total	**n (x%)**	**n (x%)**	**n (x%)**

Table 3.6: Reasons ineligible

Reason	**Frequency**	**Percent**
Total		

Table 3.7 Reported ineligible participants by randomised treatment arm with reasons

Randomised Treatment Arm				Total
Tonsillectomy		Conservative management		
Number	Reason	Number	Reason	

Section 5 Study population

Table 5.1: Baseline demographic characteristics, by treatment arm

		TRIAL		TOTAL
Variable		Treatment A N=	Treatment B N=	Total N=
Gender	F	n (x%)	n (x%)	n (x%)
	M	n (x%)	n (x%)	n (x%)

Age (years) (from randomisation log)	n Median (IQR) Range	n Median (IQR) (min, max)	n Median (IQR) (min, max)	n Median (IQR) (min, max)
Ethnic origin	**n (%)**	n (x%)	n (x%)	n (x%)
White		n (x%)	n (x%)	n (x%)
Asian (Indian/Pakistani/Bangladeshi ancestry)		n (x%)	n (x%)	n (x%)
Other Asian		n (x%)	n (x%)	n (x%)
Black or Afro-Caribbean (African or Caribbean ancestry)		n (x%)	n (x%)	n (x%)
Other ethnic origin		n (x%)	n (x%)	n (x%)
Education level	**n (%)**	n (x%)	n (x%)	n (x%)
Post-graduate		n (x%)	n (x%)	n (x%)
Degree/Professional/Vocational(e.g NVQ level 4)		n (x%)	n (x%)	n (x%)
Higher/A-level/National grade/vocational(e.g HND)		n (x%)	n (x%)	n (x%)
O-level/O grade/GCSE/Standard Grade/vocational(e.g HNC)		n (x%)	n (x%)	n (x%)
No educational qualification		n (x%)	n (x%)	n (x%)
Other		n (x%)	n (x%)	n (x%)
Employment status	**n (%)**	n (x%)	n (x%)	n (x%)
Self employed		n (x%)	n (x%)	n (x%)
Paid employment (full or part time)		n (x%)	n (x%)	n (x%)
Unemployment (actively seeking work)		n (x%)	n (x%)	n (x%)
Retired		n (x%)	n (x%)	n (x%)
Maternity leave		n (x%)	n (x%)	n (x%)
Looking after family or home		n (x%)	n (x%)	n (x%)
Full time student/at school		n (x%)	n (x%)	n (x%)
Long term sick or disabled		n (x%)	n (x%)	n (x%)
Government training scheme		n (x%)	n (x%)	n (x%)
Other		n (x%)	n (x%)	n (x%)

Table 5.2: Baseline demographic characteristics, randomised versus non-randomised participant populations

Variable		Randomised	Non-randomised
Gender	M	n (%)	n (%)
	F	n (%)	n (%)
Age (years)		Median (IQR)	Median (IQR)
Sore throat severity: Number of sore throat days in last 6 months		This information was not collected for the randomised group	Median (IQR)

6. Treatment received

Table 6.1: Tonsillectomy final summary. Shaded grey - crossovers

Randomised Arm	**Crossed over**	**Surgery received**	**Complied (surgery within 8 weeks)**	**Number (N=)**	**Additional information**
Tonsillectomy	No	Yes	Yes	n	
Tonsillectomy	No	Yes	No	n	
Tonsillectomy	No	No	No	n	

Tonsillectomy	No	No	withdrawn	n	
Tonsillectomy	Yes	No	N/A	n	
Total Tonsillectomy				**N1**	
Conservative management	Yes	Yes	N/A	n	
Conservative management	Yes	No	N/A	n	
Conservative management	No	No	N/A	n	
Conservative management	No	No	withdrawn	n	
Total conserv management				**N2**	
Trial Total				**N1+N2**	

Table 6.2: Concomitant medication reported via STAR and final 24 month follow-up visit, per-treatment group

	STAR questionnaires			Final 24 month visit			Total all sources		
	tonsillectomy	Con man	Total	tonsillectomy	Con man	Total	tonsillectomy	Con man	Total
Total reported									
Unique patients									

Withdrawals

Table 6.3: summary statistics for time from randomisation to withdrawal in weeks by arm

	Immediate Tonsillectomy	**Conservative management**	**Overall**
n	n	n	n
Median (IQR)	Median (LQ,UQ)	Median (LQ,UQ)	Median (LQ,UQ)
Range (min, max)	(min, max)	(min, max)	(min, max)

Table of line listings (appendix)

Table 6.4: MACRO line listing of AE's (n=)

Label	**Surgery date**	**AE start date**	**Days AE start after surgery**	**Severity**	**AE description**

Section 8

Table 8.1: Proportion of patients responding to weekly STAR responses, split by complete, partial and no response at increasing time points, by arm

Immediate Tonsillectomy				
Within timeframe	**Observed complete**	**Observed partial**	**Observed no returns**	**Median (IQR) number of weekly star responses**
4 weeks	n (x%)	n (x%)	n (x%)	Median (LQ,UQ)
26 weeks	n (x%)	n (x%)	n (x%)	Median (LQ,UQ)
52 weeks	n (x%)	n (x%)	n (x%)	Median (LQ,UQ)
78 weeks	n (x%)	n (x%)	n (x%)	Median (LQ,UQ)
104 weeks	n (x%)	n (x%)	n (x%)	Median (LQ,UQ)
Conservative management				

Within timeframe	Observed complete	Observed partial	Observed no returns	Median (IQR) number of weekly star responses
4 weeks	n (x%)	n (x%)	n (x%)	Median (LQ,UQ)
26 weeks	n (x%)	n (x%)	n (x%)	Median (LQ,UQ)
52 weeks	n (x%)	n (x%)	n (x%)	Median (LQ,UQ)
78 weeks	n (x%)	n (x%)	n (x%)	Median (LQ,UQ)
104 weeks	n (x%)	n (x%)	n (x%)	Median (LQ,UQ)

Table 8.2 Differential loss to follow up table

Table 8.3: Participants with zero weekly STAR responses

Table 8.4: Proportion of patients responding to weekly STAR responses, split by complete, partial and no response at increasing time points

Within timeframe	Expected	Observed complete	Observed partial	Observed no returns	Median (IQR) number of weekly star responses
6 months	N	n (%)	n (%)	n (%)	m (LQ,UQ)
12 months	N	n (%)	n (%)	n (%)	m (LQ,UQ)
18 months	N	n (%)	n (%)	n (%)	m (LQ,UQ)
24 months	N	n (%)	n (%)	n (%)	m (LQ,UQ)

Note that % is n/N*100

Table 8.5: Average time on trial (weeks) by arm for those who have made at least one weekly STAR response (N=125)

Arm		Number patients	Median (IQR) in weeks	Range
Conservative management				
Immediate tonsillectomy				
Total				

Negative binomial regression (unadjusted)

Table 8.6: Results of negative binomial regression (model 1), unadjusted, ITT group

Model	beta	SE	Test statistic	P value	95% CI (beta)	
					lower	upper
Arm: (ref = surgery) Con man						

Negative binomial regression (adjusted)

Table 8.7: Results of negative binomial regression (model 1a), adjusted for stratification factors used at randomisation (site and baseline TOI14 severity), ITT group

Model	beta	SE	Test statistic	P value	95% CI (beta)	
					lower	upper
Arm: (ref = surgery) Con man						
Site: (ref. = site x) Site 1 Site 2…						

Site n						
TOI14 baseline severity: (ref =mild) Moderate severe						
Constant						

- Similar table for analysis with continuous baseline severity

Table 8.8: Sore throat day rate over time

Randomised	**Sore throat day rate**			
	Pre surgery	x weeks after surgery	Rest of follow up	Overall
Surgery (all in surgery arm) Surgery within 8 weeks Surgery any time No surgery	NA Rate 1a Rate 2a N/A	NA Rate 1b Rate 2b N/A	NA Rate 1c Rate 2c N/A	RS all Rate 1d Rate 2d RNS1
Conservative management (all) Crossed had surgery No surgery	NA Rate 3a N/A	NA Rate 3b N/A	NA Rate 3c N/A	RC all Rate 3d RNS2
Total surgery	R2a+R3a	R2b+R3b	R2c+R3c	R2d+R3d
Total no surgery	N/A	N/A	N/A	RNS1+RNS2
Total	N/A	N/A	N/A	RS all + RC all

Section 9 secondary outcomes

Table 9.1: TOI14 completion rate and scores at time points.(% is the proportion completed out of those expected at each time point)

TOI14 (secondary outcome)	**Baseline**	**6 months**	**12 months**	**18 months**	**24 months**
Immediate Tonsillectomy n (%) Median (IQR) Mean (SD) 95% confidence internal about mean Range (min, max)					
Conservative management n (%) Median (IQR) Mean (SD) 95% confidence internal about mean Range (min, max)					
Overall n (%) Median (IQR) Mean (SD) 95% confidence internal about mean Range (min, max)					

Table 6.3: TOI 14 scores at data collection time points during trial

TO14 score		Tonsillectomy (n=XX)	Conservative management (n=XX)	Total (n=XX)
baseline	n median (IQR) Range (min max)			
6 month (post)	n median (IQR) Range (min max)			
12 month (visit)	n median (IQR) Range (min max)			
18 month (post)	n median (IQR) Range (min max)			
24 month (visit)	n median (IQR) Range (min max)			

Similar table showing TIOI14 subscales of Throat Discomfort, General Health, Resource impact, Social and psychological will also be provided.

Section 4 Form returns

Table 4.1: Participant activities completed by time point
Note^ - Allowing for window of 6 weeks for visits and form returns and 8 weeks for tonsillectomy

CRF	Number expected^	Actual number	Return rate
Baseline form (clinic visit 1)	n	n	x%
Baseline (declined patients) (section 4.2)	N/A	n	N/A
Weekly STAR responses (PRIMARY OUTCOME)	n	n	x%
Weekly STAR responses (PRIMARY OUTCOME) adjusted for weeks up to withdrawing	n	n	x%
Weekly STAR responses (PRIMARY OUTCOME) adjusted for weeks up to withdrawing excluding extra returns beyond 105 weeks	n	n	x%
Reported STAR questionnaires (based on weekly STAR responses received that are >0) Old paper STARS New electronic "Starlets" (Date implemented was 05/05/17)	n	n n n	x%
Tonsillectomy form (not accounting for withdrawals) Randomised to immediate tonsillectomy arm (not crossed over) Randomised to conservative management (crossed to surgery)	**n** n n	**n** n n	**x%** x% x%
Crossover form	N/A	n	N/A
Withdrawal form	N/A	n	N/A
Post surgery phone call1	n	n	x%
Post surgery phone call2	n	n	x%
6 month follow up form (postal) *allowing for 6 week window*	n	n	x%
12 month follow up (clinic visit 2) *allowing for 6 week window*	n	n	x%

18 month follow up form (postal) *allowing for 6 week window*	n	n	x%
24 month follow up (clinic visit 3) *allowing for 6 week window*	n	n	x%

Section 7: SAFETY ANALYSIS Adverse events

Table 7.1: reported AE's on ITT population

Related to treatment	Severity	Tonsillectomy	Conservative management
Unrelated	Mild		
	Moderate		
	Severe		
Unlikely	Mild		
	Moderate		
	Severe		
Possible	Mild		
	Moderate		
	Severe		
Probable	Mild		
	Moderate		
	Severe		
Definitely	Mild		
	Moderate		
	Severe		

Table 7.2: line listing of AE's on ITT population

Related to treatment	severity	Adverse effect	Days from last visit before AE	Days to next visit

Table 7.3: reported SAE's on ITT population

Related to treatment	Severity	Tonsillectomy	Conservative management
Unrelated	Mild		
	Moderate		
	Severe		
Unlikely	Mild		
	Moderate		
	Severe		
Possible	Mild		
	Moderate		
	Severe		
Probable	Mild		
	Moderate		
	Severe		
Definitely	Mild		
	Moderate		

Table 7.4: Adverse events categorised (n=)

AE category	**Frequency (%)**
Pain: throat or ear	n (x%)
Bleed	n (x%)
Infection/fever/temperature	n (x%)
Nausea/vomiting	n (x%)
reduced diet/swallow difficulty	n (x%)
Tiredness/Fatigue	n (x%)
Extra rows for occasion where more than one category specifies (eg pain and bleeding)	n (x%)
Other	n (x%)
Total	n (x%)

Table 7.5: line listing of SAE's on ITT population

Related to treatment	Serious Adverse effect	Days from last visit before SAE	Days to next visit
	list		

Table 7.6: Further details of the reported SAE's

Further details of reported SAE's	**Frequency (%) N=**
Subject died	n (x%)
life threatening	n (x%)
Admitted or extended hospital stay	n (x%)
Involved persistent or significant disability or incapacity	n (x%)
Other significant medical event	n (x%)
completely recovered	n (x%)