The clinical and cost effectiveness of surgical interventions for stones in the lower pole of the kidney: The **P**ercutaneous nephrolithotomy, flexible **U**reterorenoscopy and **E**xtracorporeal shockwave lithotripsy for lower pole kidney stones **R**andomised **C**ontrolled **T**rial (PUrE RCT)

SAP Version:	1.0
SAP Date:	19th October 2022
Protocol Version:	3.0
Protocol Date:	8th October 2020
ISRCTN:	98970319

0. Administration

0.1. (Co-)Sponsors

Name:	University of Aberdeen
Address:	Research and Innovation, Kings College, Regent Walk, Aberdeen AB24 3FX, UK
Name:	Grampian Health Board
Address:	Research and Development Office, Foresterhill Annexe, Foresterhill, Aberdeen AB25 2ZD, UK
Sponsor number	3-052-15

Investigators

Clinical Chief Investigator

Name:	Professor Sam McClinton NHS Grampian, Department of Urology, Aberdeen Royal Infirmary, Foresterhill, Aberdeen AB25 2ZB, UK 01224 550517 smcclinton@nhs.net

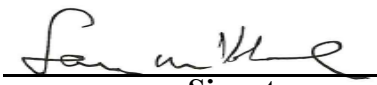
Funder

Name:	The National Institute for Health Research (NIHR), Health Technology Assessment (HTA) Programme
Funder number:	13/152/02
Funder start date:	01 December 2015

Funder end date:	28 February 2023
NIHR portfolio	188563
Trial registration: ISRCTN#	98970319
REC number:	15/NS/0113
EUDRACT number	

0.2. Signatures

By signing this document, I am confirming that I have read, understood and approve the statistical analysis plan (SAP) for the **PUrE RCT** trial.

Sam Clinton **Chief Investigator**	 **Signature**	19/10/22 **Date**
Lorna Aucott **Trial Statistician**	 **Signature**	19/10/22 **Date**

0.3. Version History

SAP version	Protocol version	Section number changed	Description of and reason for change	Date changed
Version .01	Version 1		New document	1st December 2017
Version 1	Version 3	• definite derivation of the primary outcome. Section 12.3 • Sequence and format of whole document. Added dummy tables and appendices for Proms and model codes. • Typos and minor clarification	Changed to newest SAP template in word and incorporated derivation of the primary outcome.	19th October 2022

0.4. Glossary of Abbreviations

ACCORD	Academic and Clinical Central Office for Research & Development - Joint office for The University of Edinburgh and Lothian Health Board
AE	Adverse Event
AR	Adverse Reaction
AUC	Area Under the Curve
CHaRT	Centre for Healthcare Randomised Trials
CI	Confidence Interval
CONSORT	Consolidated Standards of Reporting Trials
CRF	Case Report Form
DMC	Data Monitoring Committee
EQ-5D	EuroQol Group's 5-dimension health status questionnaire
ESWL	(Extracorporeal) shockwave lithotripsy
FURS	Flexible ureterorenoscopy
HRQoL	Health related quality of life
HSRU	Health Services Research Unit
HTA	Health Technology Assessment
ITT	Intention-to-Treat
NIHR	National Institute of Health Research
NRS	Numeric Rating Scale
PCNL	Percutaneous nephrolithotomy
PQ	Participant questionnaire
RCT	Randomised controlled trial
RCT1	RCT of FURS vs ESWL for stones with maximum dimension <=10mm
RCT2	RCT of FURS vs PCNL for stones >=10mm and <=25mm
RR	Relative risk
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SD	Standard Deviation

0.5. Contents

PUrE

This is the statistical analysis plan (SAP) and accompanying documents detailing the statistical analyses planned for the PUrE Trial. The SAP is based on the protocol (latest version here) and any deviations from the plan will be described.

1. Introduction

Renal tract stone disease is very common, with a lifetime prevalence of approximately 10% in the adult population across the world. Guidance issued by the European Association of Urology (EAU) and widely followed in UK clinical practice recommends ESWL as an option for lower pole stones ≤10mm whereas for larger stones recommended options are FURS or PCNL. However, the guidance adds that ESWL may be used for larger stones if stone factors and patient preference are favourable. Flexible ureteroscopy and laser fragmentation and PCNL are more invasive than ESWL, require a general anaesthetic, and carry a greater risk of complications.

There is some evidence to inform estimates of the relative clinical effectiveness (based upon stone free rate) of ESWL, FURS and PCNL in the treatment of lower pole stones and to guide clinical practice. However there is sparse evidence, on the impact of these treatments upon patient reported health status and quality of life outcomes (such as severity and duration of pain after intervention), their care pathway (such as the need for additional interventions) and resource use.

2. Study Aims and Objectives

The aim of the study is to determine which of FURS, PCNL and ESWL offer the best treatment outcomes in terms of clinical effectiveness and cost effectiveness for people with lower pole kidney stones seeking treatment within the UK NHS. An initial pilot phase will be built in to the trial to assess feasibility of recruitment and check appropriateness of eligibility criteria and outcome measures. The research question to be addressed is: In people requiring treatment for lower pole stones of the kidney does flexible ureterorenoscopy with laser lithotripsy result in better quality of life than standard treatment with ESWL or PCNL according to stone size, and is it cost-effective for the UK NHS?

3. General Study Design

PUrE consists of two pragmatic multicentre patient-randomised open label superiority RCTs. A summary of the trial design is shown in the figure below.

RCT 1: Flexible ureterorenoscopy with laser lithotripsy (FURS) versus (extracorporeal) shockwave lithotripsy (ESWL) recruiting patients with stones of maximum dimension ≤10mm.

The null hypothesis being tested is: the use of FURS to treat lower pole kidney stones ≤10mm will not be different to ESWL as assessed by the EQ-5D AUC up to 12 weeks post treatment.

RCT 2: Flexible ureterorenoscopy (FURS) versus percutaneous nephrolithotomy (PCNL) recruiting patients with stones of maximum dimension >10mm and ≤25mm.

The null hypothesis being tested is: the use of FURS to treat lower pole stones of the kidney >10mm and ≤25mm will not be different to PCNL as assessed by the EQ-5D AUC up to 12 weeks post treatment.

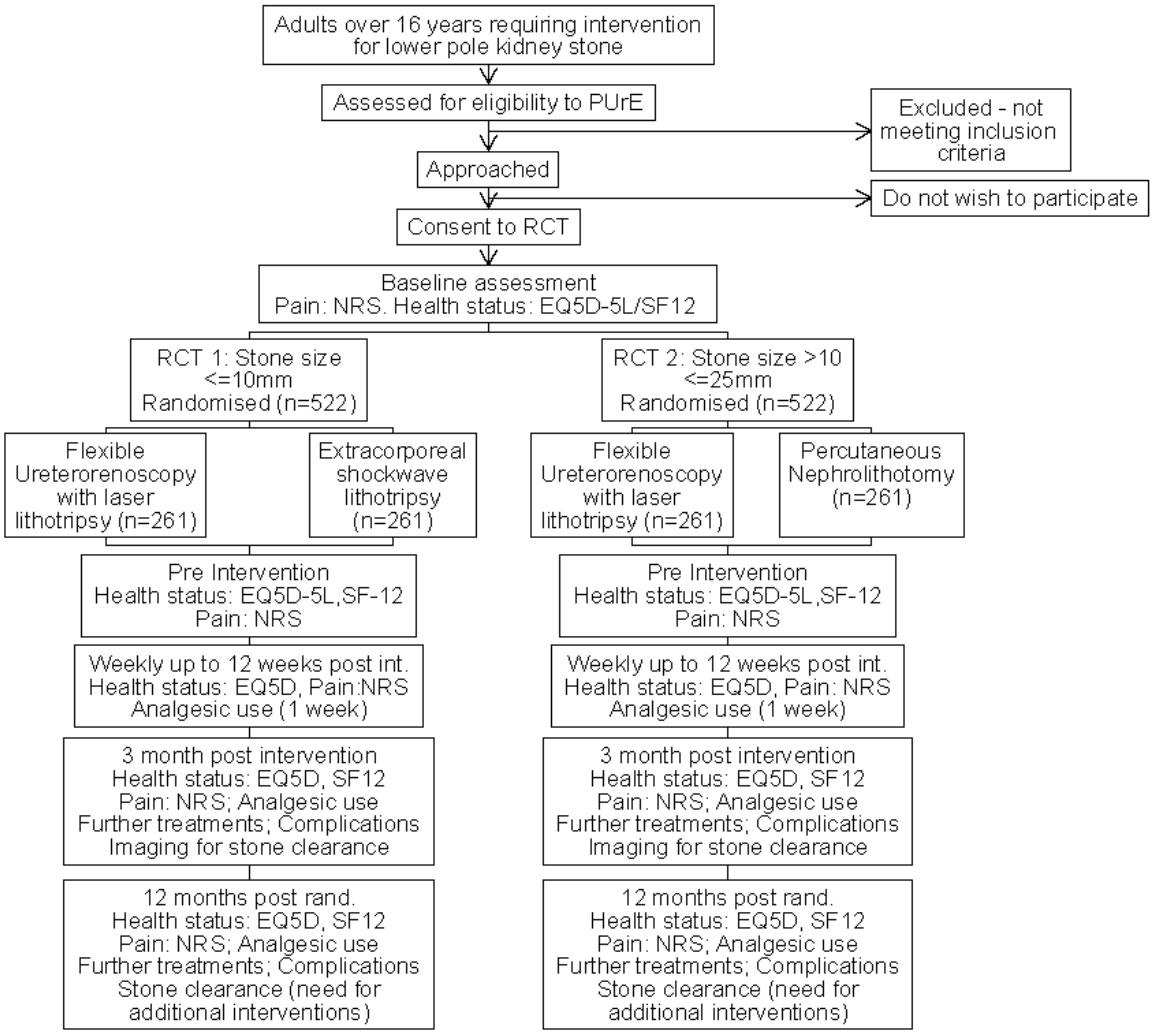


Figure 1: Trial Design

It is important to note that RCT1 and RCT2 will be analysed entirely separately, but their analysis plans are largely the same bar the treatment arm differences.

4. Interventions to be evaluated

- Experimental: Flexible ureterorenoscopy with laser lithotripsy (FURS)
- Standard: Shockwave lithotripsy (ESWL) or Percutaneous nephrolithotomy (PCNL) depending on if RCT1 or RCT2

5. Inclusion and exclusion criteria

Inclusion and exclusion criteria are recorded in detail in section 3.2.2 of the Protocol. We reproduce the basics of them here for convenience.

Inclusion criteria:

- Adults 16 years of age
- Lower pole stone≤25mm in maximum dimension with decision to treat that stone
- Presence of stone previously confirmed by CTKUB
- Able and willing to undergo either treatment for specified stone size
- Capacity to give informed consent to participate in trial which includes adherence to trial requirements

Exclusion criteria:

- Pregnancy
- Co-existing stone that takes precedence in deciding treatment modality (such as obstructing ureteric stone or large upper pole stone)
- Health or other factors that are absolute contraindications to an intervention that they may be allocated
- Unable to understand or complete trial documentation

6. Change of Status

There are several reasons why we might not be able to collect the information to fully report on randomised participants as we had expected. We try to understand what these reasons might be and hence what consequence on the study these change in status might have

Post randomised exclusions –we exclude such participants, since they were randomised before it was fully realised they were not eligible.

Participants who decline some or all of the follow-up strategies

- decline to participate and/or decline the collection of all or part of the data
- decline to have Participant questionnaire
- decline from clinical data collection
- declined treatment

Change in treatment plan

- Stone not seen/stone known to have passed
- Patient asked for conservative management
- Clinical team decided on conservative management

Unable to collect information

- Death whilst in study
- Change in address
- No responses but we have no reason

These are not mutually exclusive, with participants having possibly one or many of these states.

7. Randomisation, Allocation and Blinding

PuRE participants will be randomised using CHaRT's telephone Interactive Voice Response (IVR) randomisation application or via the web-based application. Simple randomisation is employed within each RCT. Participants with a stone ≤10mm will be randomised to either FURS or ESWL (RCT 1). Participants with a stone >10mm and ≤25mm will randomised to either FURS or PCNL (RCT 2). Randomisation programming is carried out by the CHaRT IT team, details on randomisation procedures are available at CHaRT SOP 9.9 and 10.4.

Blinding is not possible in PuRE.

8. Outcome Measures

8.1. Primary Outcome(s)

In this study, Health status is measured by the EQ-5D-5L mean score at various time points including baseline (recruitment), just prior to first intervention (or some designated time if an intervention is not formally conducted) and weekly up to 12 weeks post-intervention. In addition, responses will be collected after specific hospital events prior to the pre-intervention visit and during any post-intervention hospitalisations (eg for adverse events related to treatment).

Originally, the Primary outcome was to be the Area under the curve (AUC), incorporating baseline up to 12 weeks post-intervention. Given the potential of differing treatment waiting times and that AUC should be over a fixed time period overall we will consider one main primary outcome and a couple of other scenarios. See the primary outcome methods in section 12.3.1 and 16. Appendix A for specific details.

8.2. Secondary Outcomes

As defined in the protocol

- SF-12
- Use of analgesia
- Stone clearance by 12wk post 'intervention' (ie from date of intervention or the agree nominal date). This is categorized as complete, acceptable, or unacceptable. This is also to be reported at 12mo
- By 12 weeks, the maximum dimension of the largest fragment (if there are any) of the treated stone in mm at follow-up will be described as n/N (%) for categories <2mm, between 2-4mm, >4mm, by treatment arm, but not formally compared between treatments.
- Additional intervention (carried out or planned) at 12 weeks post-initial treatment and 12 months post-randomisation
- Treatment related complications upto12 weeks post operative only (binary y/n)

In addition the project management group a-priori were keen to investigate:

- Health status as measured by a weekly average of EQ-5D-5L from the Area under the curve (AUC), from baseline to 12 weeks post-intervention
- EQ5D -5L: 6th domain a Visual scale (VS)
- Severity of pain as measured by the NRS, (VAS) (0–10 integers) (none to unbearable)
- Pain relief over the last 7 days related to your kidney stone? (Y/N)
- Waiting times from randomisation to intervention date.

9. Timing of Outcome Measurements

Table 9-1: Timing and source of outcome measurement

Outcome	Source	B	Pre	W1-W11 †	W12 †	Additional intervention (pre and post if >12 weeks) or treatment-related hospitalisation	12M *
Health status EQ-5D-5L	PQ	✓	✓	✓	✓	✓	✓
Pain (NRS today)	PQ	✓	✓	✓	✓	✓	✓
Health profile SF12	PQ	✓	✓		✓		✓

Pain relief over last 7Days	PQ	✓	✓	✓	✓	✓	✓
Stone clearance (imaging)	CRF				✓		✓
Additional interventions received	CRF+ PQ				✓	✓	✓
Complications	CRF+ PQ				✓	✓	✓

* Post randomisation † Post intervention

10. Adverse Events

Each initial Adverse event (AE) will be considered for severity, causality or expectedness. A serious adverse event (SAE) is any AE that:

- Results in death;
- Is life threatening;
- Required hospitalisation or prolongation of existing hospitalisation;
- Results in persistent or significant disability or incapacity;
- consists of a congenital anomaly or birth defect;
- results in any other significant medical event not meeting the criteria above.

Please see the Study Protocol for more details on AEs. The number of Adverse events (AEs) and serious adverse events (SAEs) and the proportion of participants with an event will be presented. These will be tabulated and not analysed and will be summarised by Intention-to-Treat (ITT) and as treated.

11. Sample Size and Power Calculation

The primary outcome is the average AUC measured from multiple completion of the EQ-5D by each participant up to 3 months post first intervention. Good data on AUC in this patient group are sparse. In order to detect a 0.3 SD difference, with 90% power, and alpha set at 5%, 235 participants per group (470 total) are required. Such a difference in generic health status is considered clinically relevant and in terms of treatment effect size, in the small to medium range as observed in other clinical studies {Walters et al, 2005; Farivar et al, 2004}. This would equate to a difference of 0.1 in the AUC on the standardized 0-1 utility scale assuming a standard deviation of 0.33 or less. To allow for the anticipated approximately 10% of participants for whom outcome data is completely missing, and therefore the AUC cannot be calculated, it is proposed to randomise 522 participants in both RCT 1 and 2 giving a total trial population of 1044 participants.

12. Statistical Methods

Baseline demographic data will be tabulated for each group. Discrete variables will be summarised with numbers and proportions. Continuous variables by the mean and standard deviation (or median and IQR if skewed).

Any post-randomisation exclusions will be excluded from the baseline tables and all analyses.

Tables categorising the proportions of participants receiving the allocated intervention, other intervention or no intervention will be presented for each arm of RCT1 and RCT2.

12.1. General Methods

All the main analyses will be based on the Intention-to-Treat (ITT) principle. For the primary outcome(s) we will present per protocol (PP) and complier adjusted causal estimation (CACE) sensitivity analyses. Final analysis will take place after full recruitment and 12 months post randomisation or 12 weeks post-intervention whichever is the latest. The results of the trial will be presented following the standard CONSORT recommendations[3] {Consort 2020}. Baseline and follow-up data will be summarised using the appropriate descriptive statistics and graphical summaries. Treatment effects will be presented with 95% confidence intervals (CIs) (apart from subgroup analysis). There will be no adjustment to secondary outcomes for multiple testing. All eligible participants will be included in the analysis and who provided consent. Any post-randomisation exclusions will be removed and reported as such and agreed with the PMG. Model assumptions will be checked and dealt with appropriately.

Unless stated differently all Statistical significance will be based on two-sided tests with $p<0.05$ taken as the cut-off for statistical significance.

12.2. Interim Analysis

There are no planned interim analyses for efficacy or futility but an independent Data Monitoring Committee (DMC) will monitor trial progress and specifically any safety issues.

12.3. Primary Outcome

Recall that our Primary outcome is health status as measured by the EQ-5D-5L as an Area under the curve (AUC) over time. The methodology indicated in the protocol however cannot be actioned as envisioned because treatment waiting times vary not just between interventions but over the whole study period. Instead, we propose one Main Primary model and two Sensitivity Primary models that will use an area under curve but account for the full period in slightly differing ways. In addition, we propose another measure (the weekly average AUC) as a secondary outcome – see in section 12.4.

Here the primary outcome is defined as health status AUC anchored between date of pre-intervention and 12th week post-operative [#]. This AUC will be generated for each participant as described in 16. Appendix A.

12.3.1. Main model

In order to compare FURS to either ESWL in RCT1 or PCNL in RCT2, the main primary analysis model uses the AUC as describe above[#] for each participant. We will use a generalised linear model with a gaussian family and canonical link, adjusting for centres (with a robust variance structure) and health status at randomisation to accommodate quality of life at baseline thus incorporating health status over the whole period from randomisation to the 12th week post-intervention. See 17. Appendix B for specific model codes.

12.3.2. Primary Outcome - Sensitivity Model(s)

(i) This is the same as the main model but will **NOT** include the baseline health status.
(ii) This will be identical to the main model but will adjust not only for baseline health status but also waiting time (from randomisation up to the intervention date).

These models will also adjust for centres (with a robust variance structure).

12.4. Secondary Outcomes

All secondary outcomes will be similarly treated as the primary outcome(s) to test between-group change. However, where applicable time and treatment/time interactions will also be fitted as fixed effects along with baseline outcome values (as appropriate), centres (with a robust variance) and participants within centres (for the repeated measure models)

In addition to the secondary outcomes specified in the protocol, we plan to assess:

- an alternative to the main primary outcome model where the area under the curve for EQ5D 5L will be anchored at randomisation. To account for the varying waiting treatment times, we propose to standardise by dividing by the number of weeks this involves and hence using the weekly average health status AUC ($WAAUC_{hs}$).
- Also added is a comparison of mean waiting times from randomisation to intervention between arms by treatment arms.

All outcomes are to be assessed on a superiority basis.

See 18 Appendix C for the model suggested Stata codes.

12.5. Subgroup Analyses

Subgroup analyses will explore the possible modification of treatment outcome effect by important factors on the Main Primary outcome model, its two sensitivity models and for the secondary outcome, Additional intervention as determined at 12 months (either completed or still planned). The subgroups identified are:

- participant body mass index: This measure may not be complete for all participants since it is mainly collected at the time of intervention (not all participants have an intervention eg the stone passed, conservative management was agreed etc…). Where possible, the patient records will be examined for cases where this may have been observed at other times. Hence, we will compare the possible impact on treatment and

a participant weight being Normal/ Overweight/ Obese/ Extreme, as a casewise analysis.

- stone size (maximum dimension and volume). Since RCT1 only includes those with initial stones <10mm we will only consider this for RCT2. In addition, only the maximum dimension will be assessed, comparing 10-15mm vs >15mm.
- stone density on CTKUB (Hounsfield units) to compare <=1000 vs >1000 .
  skin to stone distance. This derived measure is not routinely considered and so while the sub-group has been decided a-priory, establishing the cut-off categories will be a post-hoc process in collaboration with our Project management team.

A stricter level of statistical significance (2-sided 1% significance level) will be applied to these analyses given their exploratory nature. Corresponding 99% confidence intervals will therefore be calculated.

See 22 Appendix G for specific details of deriving the stone related subgroups. All subgroups will be conducted by including treatment-by-factor interactions in the model see 19 Appendix D for specific code.

12.6. Technical Factors impact

We will also describe within each allocated group how technical factors might impact the main the Primary outcome (and its associated sensitivity outcomes) and Additional intervention as determined at 12 months (either completed or still planned). This will be when these technical factors have been applied at both the initial intervention and also for any 2nd intervention permitted within the treatment pathway.

- access sheath versus no access sheath and digital versus non digital instrument (FURS)
- fixed site versus mobile device (ESWL);
- caliber of access track (PCNL).

12.7. Other Sensitivity analysis

For the Main Primary Outcome model (and its associated Sensitivity models) we will also conduct:

- Multiple imputation of missing EQ-5D outcome data (at any timepoint) will be used in sensitivity analyses following the strategy outlined in {White et al 2011}
- Implement these into the Mixed model to derive a summary AUC effect for each intervention group as outlined in {Bell et al, 2014} for both the original from randomisation and from pre-intervention

12.8. Compliance

We will explore the influence of compliance on the treatment effect for the primary outcome by doing a per-protocol analysis and complier adjusted causal estimation (CACE) using instrumental variable regression {Little RJ and Rubin DB, 2000}.

12.9. Missing Data

12.9.1. Missing Outcome Data

There should be no missing primary outcome data. None-the-less the sensitivities of treatment effect estimate to missing outcome data will be explored; these models will explore the robustness of the treatment estimate to whatever small amount of missing data there is. We will follow the strategy outlined in {White et al 2011}. The analysis will use all available data that we believe are valid under the assumption of missing at random. We will then use a suite of sensitivity analysis to explore the robustness of the primary outcome to departures from assumption, including all randomised participants. If required, sensitivity analyses will include multiple imputation, and imputing a range of values for missing data under missing not at random assumptions.

12.9.2. Missing Baseline Data

Data missing at baseline will be reported as such. If required, primary and/or secondary outcome data will be imputed with centre specific mean for continuous data and missing binary/categorical data will include a missing indicator {Sullivan et al, 2018}.

12.10. Statistical software

All statistical analyses will use Stata (vs 17). All results will be processed directly into PDF/Word from Stata via the use of putdocx commands for the final Statistical Report.

12.11. Derived variables

There are several derived variables, including the basic elements of the primary outcome. These have all been collected using validated questionnaires which require scores and/or subscales to be calculated. Codes for these are developed in-house (unless externally provided eg EQ5D-5L and SF-12), checked and the code verified using dummy data by an independent statistician. See 21. Appendix F for specific details relating to agreed methods and methods of allowing missingness (with possible imputation methods).

13. COVID-19

The effect of COVD-19 will be explored. In the first instance, periods before, during and after COVID-19 will be summarised using appropriate descriptive statistics and graphical summaries. If need be, formal analysis will be carried out to explore the effect of COVID-19.

14. REFERENCES

Bell, Melanie L.; King, Madeleine T. and Fairclough, Diane L. Bias in area under the curve for longitudinal clinical trials with missing patient reported outcome data. SAGE Open, 4(2), 2014.

Consort 2010 Consort - CONSORT 2010 (consort-statement.org)

Farivar, Sepideh S.; Liu Honghu, and Hays, Ron D. Half standard deviation estimate of the minimally important difference in hrqol scores. Expert review of pharmacoeconomics and Outcomes Research, 4, 2004.

Little RJ and Rubin DB. Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. Annu Rev Public Health 2000

Sullivan, Thomas, R; White, Ian R; Salter, Amy B; et al. Should multiple imputation be the method of choice for handling missing data in randomized trials? Statistical Methods in Medical Research, Vol 27, Issue 9, 2018
https://journals.sagepub.com/doi/10.1177/0962280216683570

White IR, Horton NJ, Carpenter J, Pocock SJ. Strategy for intention to treat analysis in randomised trials with missing outcome data. *BMJ* 2011; 342:d40.

Walters, Stephen J. and Brazier, John E. Comparison of the minimally important diffrence for two health state utility measures: Eq-5d and sf-6d. Quality of Life Research, 14, 8 2005.

15. Dummy Tables

- Treatment 1 is the experimental intervention i.e. FURs
- Treatment 2 is the standard intervention and will be ESWL for RCT1 or PNCL for RCT2

15.1. Baseline demographics

Table 15-1: Baseline table

	Treatment 1	Treatment 2
Gender †		
male		
female		
Level of pain today *		
Had stone pain in last 7 days †		
yes		
no		
No. of days taken pain relief in last 7 days *		
Age(years) *		
Stone size *		
dimension 1 (D1)		
dimension 2 (D2)		
craniocaudal (D3)		
Hounsfield Unit *		
Skin to stone distance *		
0 degrees		
45 degrees		
90 degrees		
Previously had kidney stony needing treatment†		
EQ5D *§		
EQ5D VAS *		
SF12: *		
physical component score		
mental component score		
BMI at preintervention ‡		

*Continuous variables (pseudo) n, mean (sd) median,IQR, (min, max)
§ Derived using NICE recommendations {https://euroqol.org/support/analysis-tools/cross-walk/.} that accounts for age and sex
†Categorical/ordinal N n (%)
‡ mainly collected at time of intervention – some did not have a formal intervention due to it being known or assumed the stone had passed, medical advice changed to conservative management or by patient decision – were possible a measure will be retrieved from the patient records during the study period.

15.2. Timepoint summary descriptions of QoL measures and Pain

Table 15-2: Summary descriptions of QoL measures and Pain for each time reported

Timepoint XX	**Treatment 1**	**Treatment 2**
EQ-5D - mean (sd)		
EQ-5D Visual scale - mean (sd)		
SF-12:		
physical component score		
mental component score		
Pain (NRS): level of pain today - median (IQR)		
No. of days taken pain relief in last 7 days - median (IQR)		
Pain related to kidney stone during last 7 days – n(%)		
Yes		
No		
missing		
Stone clearance n (%) at 12 weeks		
Cleared		
Acceptable		
Unacceptable		
Cleared after 1st intervention		
Cleared after 2nd intervention		
Largest dimension remaining of the treated stone at 12 weeks		
Not reported: no image		
None present		
<2mm		
2-4mm		
>4mm		

These data will only represent those who responded to the XX-timepoint participant questionnaire.

XX represents time points at pre-intervention date, weeks 1-12 post intervention and 12 months post randomisation were available.

Where there is more than one form (eg an additional/supplementary) in any one week, then their average will be presented. In the case of Pain (NRS) if any one of them indicate pain in that week this will be set as 'yes'.

As a result of possible multiple arms the number (and proportion) will also be presented

15.3. Treatment Switching proportions

All results are Binary and so will be reported as N n (%)

Table 15-3: Treatment Received vs Randomised allocated treatment

	Randomised to Treatment 1	**Randomised to Treatment 2**
Received Treatment 1		
Received Treatment 2		
Received other treatment		
Declines treatment		
Died before treatment		
Had a complication		

15.4. Complications by Treatment

Table 15-4.1: Treatment Received vs Allocated Randomised Intervention and those who had short-term complications 12 weeks post intervention

	Randomised to Treatment 1	**Randomised to Treatment 2**
Received Treatment 1		
Received Treatment 2		
Received other treatment		
Declines treatment		
Died before treatment		
Had a complication		

Table 15-4.2:Treatment Received vs Allocated Randomised Intervention and those who had longer-term complications 12 months post randomisation

	Randomised to Treatment 1	**Randomised to Treatment 2**
Received Treatment 1		
Received Treatment 2		
Received other treatment		
Declines treatment		
Died before treatment		
Had a complication		

15.5. Safety

All results are Binary and so will be reported as N n (%)

Table 15-5: Trial Safety by 12 weeks post-intervention

	Treatment 1	Treatment 2
Clavien grade 3a		
Clavien grade 3b		
Clavien grade 4a		
Clavien grade 4b		
Clavien grade 5		

15.6. Primary outcome models

Table 15-6: Primary outcome - treatment comparison models (main and two sensitivities)

	Treatment 1	Treatment 2	Effect size (95% CI)	p-value
ITT				
Main- AUC over the 12 weeks from pre-intervention with Baseline EQ5D 5L as main covariate and any cluster effect accounted for by a robust variance on Centre				
Sensitivity 1- AUC as above with Baseline EQ5D 5L as main covariate and Waiting time				
Sensitivity 2- AUC as Main but without Baseline EQ5D 5L as covariate				

The derivation of AUC's are described in 16. Appendix A. Any missing baseline covariates will be imputed, also defined in this appendix. See 17. Appendix B for details of analyses and code.

***Similarly tables for Per Protocol (PP) and Compliance (CACE) sensitivity analyses.

15.7. Secondary outcomes comparisons

Table 15-7: Secondary outcomes - treatment comparison models

	Treatment 1	**Treatment 2**	Effect size (95% CI)	p-value
WAAUC –weekly average of AUC anchored at baseline to 12 weeks post-intervention				
Pain (NRS): model for all timepoints				
Base (ref)				
W1				
⋮				
W12				
SF-12:				
physical component score				
Base (ref)				
12 weeks post intervention				
12 months post randomisation				
mental component score				
Base (ref)				
12 weeks post intervention				
12 months post randomisation				
Use of analgesia: model for all timepoints				
Base (ref)				
W1				
⋮				
W12				
At 12 weeks[a]				
Cleared				
Acceptable				
Unacceptable				
At 12 months [b]				
Cleared				
Acceptable				
Unacceptable				
Complications (CD>3) y/n) n (%)				
After initial intervention				
12w[a]				
12mo[b]				
After second intervention (if done),				
12w[a]				
12mo[b]				
Additional intervention by 12 weeks [a]				
Additional intervention by 12 months [b c]				

[a] 12 weeks post intervention
[b] 12 months post randomisation
[c] Additional intervention carried out or still planned
[d] or nominated date if there was no formal intervention

See 18. Appendix C for details of analyses and code

Table 15-8: Additional Intervention summary summaries by 12 weeks

	Treatment 1	**Treatment 2**	Effect size (95% CI)	p-value
Interventions at person level				
Interventions total				
Types of intervention:				
ESWL				
FURS				
PCNL				
Other				

Table 15-9: Additional Intervention summaries by 12 months

	Treatment 1	**Treatment 2**	Effect size (95% CI)	p-value
Interventions at person level				
Interventions total				
Types of intervention:				
ESWL				
FURS				
PCNL				
Other				

See 20. Appendix C for Additional Intervention definitions and allowed treatment pathways

15.8. Subgroup Analyses:

Table 15-10a: Main Primary Outcome - Treatment effect account for subgroups

	Treatment effect (95% CI)	Difference in treatment effects (95% CI)	p-value
Subgroup 1			
Subgroup 2 ….			

Table 15-10bi: Sensitivity 1 Primary Outcome - Treatment effect account for subgroups

	Treatment effect (95% CI)	Difference in treatment effects (95% CI)	p-value
Subgroup 1			
Subgroup 2 ….			

Table 15-10bii: Sensitivity 2 Primary Outcome - Treatment effect account for subgroups

	Treatment effect (95% CI)	Difference in treatment effects (95% CI)	p-value
Subgroup 1			
Subgroup 2			
….			

See 19. Appendix D for details of analyses and code

15.9. Technical Factors summaries:

Recall each of the technical factors impact just one of the treatment arms in both RCT1 and RCT2.

- TF1- access sheath versus no access sheath and digital versus non digital instrument (FURS) - RCT1 and RCT2
- TF2a- fixed site versus mobile device (ESWL) - RCT1 only
- TF2b- caliber of access track (PCNL) – RCT2 only

The following tables lay out the summaries for each outcome in each treatment arm for each Technical factor (specific to that treatment arm in each of RCT1 and RCT2)

Table 15-11i: Treatment descriptive summaries of Primary Outcome (main and associated sensitivities) and Additional intervention, relating to the technical factors (TF as specified above) – TF used for **1st intervention**

	Treatment	Mean (SD) (95% CI)
Main Primary Outcome		
TF1	FURS	
TF2a/b	ESWL/PNCL	
Sensitivity i Primary Outcome		
TF1	FURS	
TF2a/b	ESWL/PNCL	
Sensitivity ii Primary Outcome		
TF1	FURS	
TF2a/b	ESWL/PNCL	
Additional intervention*		
TF1	FURS	
TF2a/b	ESWL/PNCL	

*****12 months (either completed or still planned)

Table 15-12ii: Treatment descriptive summaries of Primary Outcome (main and associated sensitivities) and Additional intervention, relating to the technical factors (TF as specified above) – TF used for **2nd Intervention** (if permitted within the treatment pathway).

	Treatment	Mean (SD) (95% CI)
Main Primary Outcome		
TF1	FURS	
TF2a/b	ESWL/PNCL	
Sensitivity i Primary Outcome		
TF1	FURS	
TF2a/b	ESWL/PNCL	
Sensitivity ii Primary Outcome		
TF1	FURS	
TF2a/b	ESWL/PNCL	
Additional intervention*		
TF1	FURS	
TF2a/b	ESWL/PNCL	

*****12 months (either completed or still planned)

Dummy Figures

RCT1: for stones ≤10mm: Treatment 1 is FURS and Treatment 2 is ESWL

RCT2: for stones>10mm: Treatment 1 is FURS and Treatment 2 is PCNL

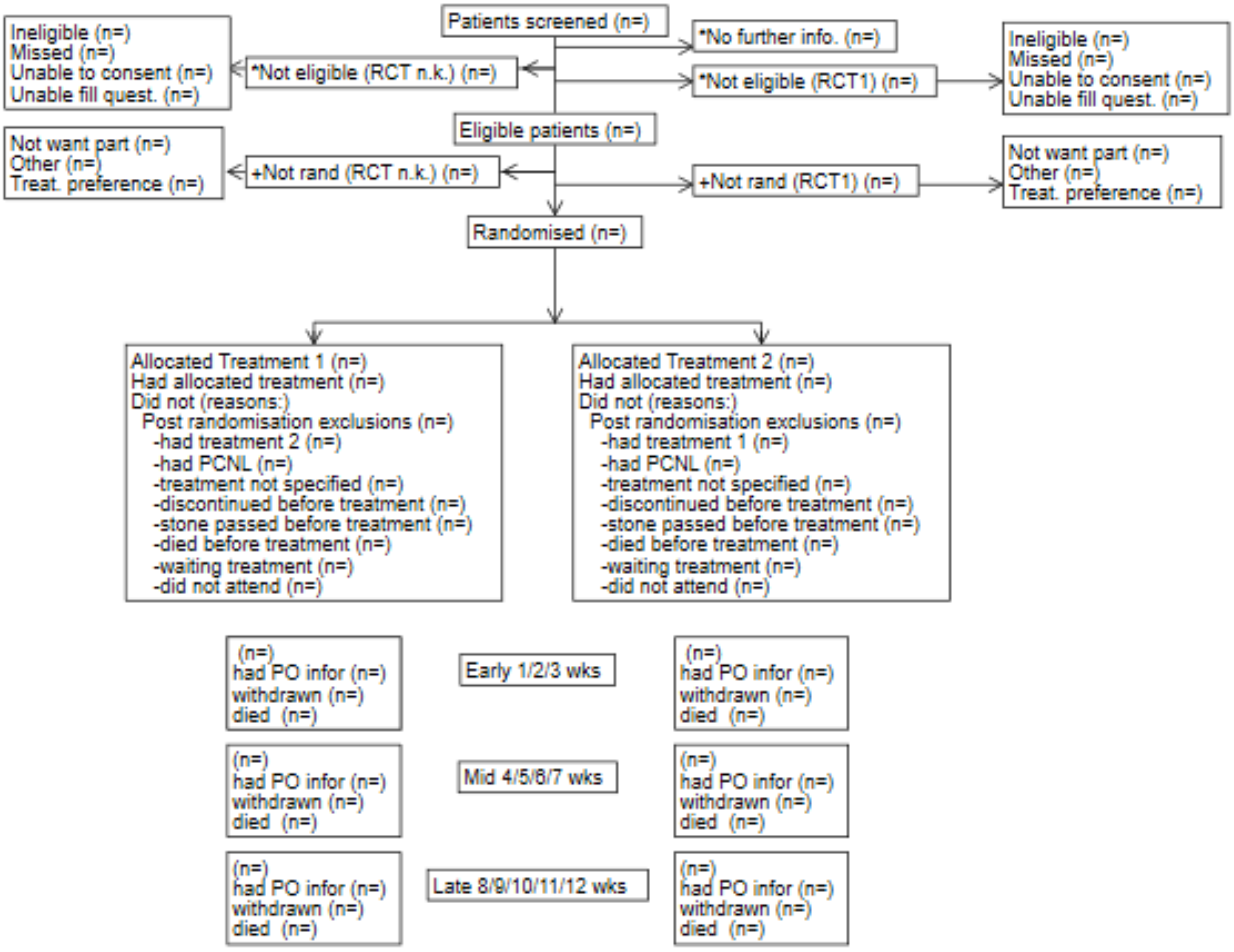


Figure 2: CONSORT diagram format for both RCT1 and RCT2

16. Appendix A: Derivation of the EQ-5D-5L Area under the curve (AUC)

The primary outcome is health status as measured by the EQ-5D-5L AUC reported from an anchor point (see 17 Appendix B:) up to and including 12 weeks post-intervention.

Using the trapezoidal area formula, a section of the AUC of the EQ-5D between two consecutive follow-up observations can be calculated as the mean of the two EQ-5D values multiplied by the time difference in between. Thus, for each time epoch, an element of the area under the curve will be determined using the formula

$$a_f = 0.5\ (e_f + e_{f-1})(t_f - [t_{f-1}]), \text{ over}$$

where t is the time of a follow-up observation f since randomisation, e_f is the EQ-5D collected at that f^{th} 'follow-up' point.

To find the total AUC, all these time epoch elements will be cumulated using Σa_f

A participants will be included if the following criteria hold: -

(i) Complete EQ-5D at baseline and/or pre-intervention (see Notes 1 and 2 below on missing data),

(ii) Complete EQ-5D observations for at least one or more occasions between 1-3 weeks post intervention

And

(iii) Complete EQ-5D for at least 1 questionnaire from weeks 4-12 week post intervention

Notes:

1. Missing baseline values will be imputed for the EQ5D index using the Centre mean average within each treatment arm. Although this missingness is expected to be very low, this approach will ensure that a participant who had made sufficient contribution otherwise is not excluded for the main model which requires the baseline as a covariate.
2. Pre-intervention EQ-5D may have more missingness. Provided sufficient repeated measures for criteria ii-iii above, we will consider imputing the nearest previous observation carried forward even if this is at baseline. Examination of the relationship between those with baseline and pre-intervention (and any other points between) will help validate this approach or not.
3. If the last valid observed value is >6 weeks post-intervention but <= 12 weeks, then the 12th week will be imputed as the last value carried forward.
4. If the last observation is >3 weeks post-intervention but less than 6 weeks, then other outcomes (eg death, complications, additional interventions other than those in the accepted treatment pathway) will be examined. If considered stable this last observed value will be considered sufficient to be imputed as the last value carried forward for the 12th week.
5. Any other missing EQ-5D for post intervention follow up will remain missing for the main ITT analysis. However, a sensitivity analysis will use multiple imputation methods to estimate such follow-up data points appropriate to the missing time since intervention if there are no observations in one or more of the time blocks (i)-(iii) described above. The EQ-5D mean score will be derived as above[§]. These will

originally be imputed at item level (truncated for each domain score 0-5) and then the EQ5D QoL index derived according to the rules of the EuroQol EQ-5D 5L instrument

6. Negative EQ-5D indexes will be taken as 'negative'. This will decrease the total calculated AUC reflecting a worse Quality of Life overall as is required.

17. Appendix B: Primary outcome model Stata codes

a. Main model Stata Code: Let *Y* be the AUC as defined in 16. Appendix A. For the Main Primary Outcome model this will be the AUC using all observations from intervention up to and including 12 weeks post-intervention. To ensure baseline is taken into account, the baseline value (*Eq_0*) will be adjusted as well as the potential effect for Centres (*C)* using a robust cluster variance. Treatment arms are represented by *Tx=1* (for FURs -intervention) or *Tx=0* for the standard i.e ESWL (RCT1) or PNCL (RCT2). Although RCT1 and RCT2 will be assessed separately they will follow the same analyses.

glm Y Eq_0 Tx, robust vce(cluster C) level(95) nolog

b. Two additional Sensitivity Primary model Stata codes

(i) *glm Y Tx, robust vce(cluster C) level(95) nolog*

(ii) glm Y Eq_0 WaitingTime Tx , robust vce(cluster C) level(95) nolog

18. Appendix C: Secondary outcome model codes

These will be similarly analysed to the Primary outcome(s) using appropriate models (see table below) dependent on outcome data type. Adjustments will be made for the outcome at baseline (when appropriate see below), treatment*time interaction (if applicable) as fixed effects. The *C* Centres will be adjusted using a robust variance and participants *P* within each centre will be treated as a random effect for measures analysed over time. *Y* represents each specific secondary variable and *Y_0* its baseline measure, although some measures are only post intervention. In our treatment arm comparison models with repeated measure, time is represented by nominal points *t=0, 1,…,12* where *0* represents baseline and thereafter 1-12 represents the observation mostly closely aligned to the weeks 1-12 post treatment (or 'nominated point').

Table 18-1: Code to illustrate model with specific family and link functions for each secondary outcome

Outcome *Y (Y_0* is the value at baseline when avaliable*)*	Model code	*Y_0*
(1st) Complications categorised as y/n by Clavien grade >3 or not: Separate models for: • After initial intervention ○ *At 12 weeks* ○ *At 12 months* • After second intervention,	*glm Y Tx, family(poisson) link(log) robust vce(cluster, C) level(95) eform* *we will extract the RR and Adjusted risk difference found by using the post-hoc command margin r.Tx*	no

○ *At 12 weeks* ○ *At 12 months*		
(2nd) Need for additional intervention (y/n) at • At 12 weeks 'Completed' • At 12 months 'Completed/still planned'	*glm Y Tx, family(poisson) link(log) robust vce(cluster, C) level(95) eform* *we will extract the RR and Adjusted risk difference found by using the post-hoc command margin r.Tx*	no
(3rd) Stone clearance: defined as complete, acceptable or unacceptable Separate models for: • After initial intervention (or 'nominated' time point) *Up to …,* • At 12 weeks post intervention (or 'nominated' time point)	*glm Y Tx, family(ordinal) link(logit) vce(cluster, C) level(95) eform*	no
Severity of pain (NRS): measure the level of pain today All timepoints and additional times…	If outcome is normal => me*glm Y Tx##i.t Y_0, reml robust vce(cluster, C) level(95) nolog || P:*	yes
SF-12: Baseline as reference then at • 12 weeks post intervention ○ Have a treatment (or nominated) date ○ Still waiting treatment • 12 months post randomisation ○ Have a treatment (or nominated) date ○ Still waiting treatment	*meglm Y Tx Y_0, reml robust vce(cluster, C) level(95) nolog*	yes
Use of analgesia: Measured by 'how many days was pain relief used over the last 7 days?' at all timepoints and any additional times (but aligned to nearest week)	If outcome is normal *mixed Y Tx##i.t Y_0, reml robust vce(robust) level(95) nolog || P:*	yes
Largest dimension of the treated stone: Only monitored during follow up for any hospital visit [mapped to nearest weekly visit] (in tblCRFFollowUp) as <2mm, 2-4mm, >4mm	outcome is ordinal => *glm Y Y_0 Tx##i.t, robust vce(cluster, C) family(ordinal) link(logit) level(95 eform || P:*	Yes - the largest of D1/D2/D3 measured at baseline (mm)
A-priori but additional to the Protocol		
Weekly Average AUC (WAAUC) ie anchored at baseline through to the time of the intervention and on up to 12 weeks post intervention	glm *Y Tx [var1… varp] robust vce(cluster C) level(95) nolog || P:*	no
Treatment waiting times	*glm Y Tx, robust vce(cluster, C) level(95)*	no

For models with time interactions, the estimates are extracted using

xlincom (_b[1.Tx] + _b[1.Tx #2.t]) … (_b[1.Tx] + _b[1.Tx #12.t]) ///
(_b[2.Tx] + _b[2.Tx#2.t]) … (_b[2.Tx] + _b[2.Tx #12.t]), level(95)

19. Appendix D: Planned subgroup model codes

These will be identical to the original Outcome model but with an added treatment*subgroup interactive term ie

a. Main Primary outcome model with subgroups
   glm Y Eq_0 + Tx ##S_k, reml robust vce(cluster C) level(95) nolog

b. Sensitivity Primary outcome model(s) with subgroups
   (i) *glm Y Tx##S_k, reml robust vce(cluster C) level(95) nolog*
   (ii) *glm Y Eq_0 WaitingTime Tx##S_k, reml robust vce(cluster C) level(95) nolog*

where k=1,…, l and l represents the total number of groups in each particular subgroup (S).

20. Appendix E: Additional intervention- definitions

Additional Intervention is any procedure outside of the Treatment Pathways defined for this trial. Classification of the Treatment pathway varies slightly for each of the interventions proposed for each treatment arm (of both RCT1 and RCT2)

Table 20-1: Permitted Treatment Pathways for each treatment arm in both RCT1 and RCT2

Treatment Arm	Treatment Pathway (defined for the PUrE trial)
FURS is a surgical procedure where a flexible endoscope is passed into the kidney through which laser energy is used to fragment the stone.	Expect there to be a single procedure although, an additional FURs procedure will be considered provided it takes place with six weeks of the initial FURS procedure. NB: • Placement or removal of a stent or catheter at time of surgery is accepted • Insertion of stent/catheter/neph post randomization but pre-intervention will also be classed as additional intervention
ESWL, involves a (sound) shock-wave, outside the body focused onto the kidney stone through the patients flank skin, causing it to fragment	two separate ESWL treatments will be considered as part of the initial ESWL intervention strategy if within an eight week period NB: • Placement or removal of a stent or catheter at time of surgery is seen as additional intervention • Insertion of stent/catheter/neph post randomization but pre-intervention will also be classed as additional intervention
PCNL is a surgical procedure to remove stones via a small (10mm) incision in the patients flank. It uses a needle, contrast dye, the use of a hollow rigid access sheath and a rigid metal telescope (nephroscope) to eventually see the stone and either retrieve it whole or to fragment it using a variety of energy delivery	a single PCNL treatment is expected to be required to completely remove stones up to 25 mm. NB: • Placement or removal of a stent or catheter at time of surgery is accepted

devices. In addition a urinary catheter may be inserted to drain the bladder for a short period after the procedure.	• Nephrostomy also part of the treatment plan if done at time of Surgery – both insert and removal • Insertion of stent/catheter/neph post randomization but pre-intervention will also be classed as additional intervention
No treatment – agreed… stone (may) has passed/asymptomatic/other reasons	If some intervention treatment is carried will be classed as additional intervention NB: Insertion of stent/catheter/neph post randomization but pre the nominated 'intervention date' will be classed as additional intervention

21. Appendix F: Description of PROMS and details of how missing values are handled for each

EQ5D-5L Score all timepoints and supplementary/ additional	The validated Stata code from EuroQol crosswalk to the 3L tool maybe found for United Kingdom using: https://euroqol.org/support/analysis-tools/cross-walk/. This is required currently by NICE. Any missing elements then the index score is classed as missing
Numeric Rating scale (NRS) all timepoints and supplementary/additional	The NRS is a **segmented numeric version of the visual analog scale (VAS)** in which a respondent selects a whole number (0–10 integers) that best reflects the intensity of their pain. If missing this is only to be imputed if all the data it is
SF-12 (baseline, 12 weeks & 12 months)	In house validated code
Stone Clearance **complete clearance** of the target stone from the urinary tract defined as no further action or observation required for that stone; **acceptable clearance** where observation is required but no intervention planned; and **unacceptable clearance** where further intervention will be required. At: 1. Up to 12week post treatment	From the 8-12 week post randomisation of follow-up CRF- Based on Q2. Define Complete as yes(1) if: • Image done No(2) & stone passed ticked & FT No(2) • Image done Yes(1) & Stone present No(2) & FT No(2) • image done Yes(1) & Stone present No(2) & fragments No(2) & FT No(2) • **acceptable** • image Yes(1) Stone present yes(1) & fragments Yes(1)& largest <2mm & FT No(2)) • Image done No(2) & FT No(2) • image Yes(1) Stone present yes(1) & fragments Yes(1)& largest >2mm & FT No(2) **unacceptable** • image Yes(1) Stone present yes(1) & fragments No(2)& FT No(2) • Image No(2) & FT Yes(1) • image Yes(1) & FT Yes (2) if this is missing we class it as missing

and 2. 12 months post randomisation	IF FT is 'yes' at this point, do we will specify technique descriptively as proportions %? NB at 12 months we only ask is the stone cleared and is there need for additional intervention…ie no image • if Q1 Y and Q2 all ALL y -> cleared • if Q1 N and Q2 No and No -> acceptable • if Q1 N and Q2 No and yes -> unacceptable "Additional" forms may also indicate stone clearance- **To be checked**! If missing, we class it as missing

22. Appendix G: Determination of subgroup variables our a-priory categories for each and details of how missing values are handled for each

We have collected a number of characteristics about each participants the index stone, (stone dimensions, skin to stone distance, Hounsfield units). These need to be described at baseline. In addition, for the subgroup analysis for each we will include an interaction into the original primary outcome model to reflect the effect of Subgroup categories by treatment arm interaction.

Any missingness in these variable groups will be described, however for the subgroup analyses any missing data will be treated as case-wise analysis.

1. **Stone dimensions:** We collect 3 stone dimensions: 2 maximum points (D1 and D2) and the craniocualdal length (D3) all in mm, via imaging.
    a. The volume of a sphere is just

$$v = \frac{4\,\pi r^3}{3}$$

which in this example may be estimated by our three 'diameters' as

Vol= (D1 * D2 * D3) * π / 6

Missingness – if any one of these is missing, volume will be missing. For trials it is sufficient to describe this but then for analyses impute baseline.

2. **Skin to stone distance:** We have collected Skin to Stone Distance SSD at the 0, 45 and 90 degrees, in mm. We will use the mean of ALL three.

3. **Hounsfield Unit:** We will summarise and describe the Unit, reporting it's mean (sd) or median (IQR: min, max) if skewed.

23. Appendix H: Full details of EQ5D 5L summary information collected at each timepoint

	Treatment 1															Treatment 2														
Timepoint	b	p	w1	w2	w3	w4	w5	w6	w7	w8	w9	w10	w11	w12	12m	b	p	w1	w2	w3	w4	w5	w6	w7	w8	w9	w10	w11	w12	12m
Mobility n (%)																														
No problem																														
Some problems																														
Extreme problems																														
missing																														
Self-care n (%)																														
No problem																														
Some problems																														
Extreme problems																														
missing																														
Usual activities n (%)																														
No problem																														
Some problems																														
Extreme problems																														
missing																														
Pain/discomfort n (%)																														
No problem																														
Some problems																														
Extreme problems																														
missing																														
Anxiety/depression n (%)																														
No problem																														
Some problems																														
Extreme problems																														

b: Baseline measure; p: pre-intervention; wx: week x (x=1 to 12) ; 12m: 12 months