Clarifying the management of men with recurrent urethral stricture: a pragmatic, randomised, multicentre superiority trial of **op**en urethroplasty versus **en**doscopic urethrotomy.

www.opentrial.co.uk

Statistical Analysis Plan

ISRCTN: 56465715

Graeme MacLennan (Senior statistician):

Signature:
Date: 17/10/2017

Robert Pickard (Chief Investigator):

Signature:
Date: 29th September 2017

VERSION 2.0

September, 2017

Contents

1. Objective

The primary objective is to determine the relative clinical effectiveness and cost-effectiveness of open urethroplasty against the standard of endoscopic urethrotomy for the treatment of men with recurrent bulbar urethral stricture within the NHS. Clinical effectiveness will be assessed by symptom control over 24 months.

2. Study design

This is a multi-centre, pragmatic patient randomised two-arm superiority trial comparing open urethroplasty (experimental) against endoscopic urethrotomy (control) for men with recurrent bulbar urethral stricture. The trial is set in a range of specialist and general NHS urology units.

3. Statistical principles

3.1. Randomisation

Eligible and consenting participants will be randomised to one of the two intervention groups using the proven 24-hour telephone Interactive Voice Response (IVR) randomisation application or via the web-based application, both hosted by CHaRT. The randomisation algorithm will use recruitment site and time since last procedure (< 12 months or ≥ 12 months) as minimisation covariates to allocate treatment to intervention and control groups in a 1:1 ratio. A random element will be incorporated into the randomisation algorithm.

Assignment to either urethroplasty or urethrotomy will not be blinded to either the participant or investigator or the local research staff (non-blinded study). However central trial staff responsible for data management, entry and analysis will be blinded to allocated intervention where possible.

3.2. Sample size

The plan is to recruit 500 participants to the study. In order to detect a 0.3 SD difference with 90% power (2-sided 5% significance level), 235 participants per group (470 in total) are required. This would equate to being able to detect at least a 0.1 difference in the AUC on the standardized 0-1 utility scale, assuming a SD of 0.33 or less. The SD of the ICIQ-MaleSF symptom AUC in a previous study was 0.15[2] . A larger SD has been conservatively allowed for in recognition of the more representative population to be recruited to this trial and the shorter follow-up period in the previous study. Such a difference in symptom burden and associated quality of life has been observed in different clinical areas for health-related quality of life (HRQoL) measures[4]. In terms of treatment effect size, this is in

the small to medium range as observed in other clinical studies[5]. To allow for the anticipated approximately 5% of participants for whom outcome data is completely missing, and therefore the AUC cannot be calculated, it is proposed to randomise 500 participants. Based on findings from recruitment to the ProtecT trial which randomises between surgery and less invasive options for men with localised prostate cancer, we conservatively estimate a 55% agreement to participate rate amongst those eligible requiring 910 men to be approached[6].

Sample size update

Assessment of ongoing recruitment rates showed the estimated sample size was unlikely to be feasible in a fundable time period, resulting in a reassessment of the original sample size calculations. Three parameters informed calculations – the minimum clinically important difference (MID) for the primary outcome, power, and the assumed standard deviation of the primary outcome. We felt that the minimum clinically important difference (0.1) had been established by patient and clinician consensus and should therefore be maintained. Reduction of the required power from 90% to 80% would be a possibility but may prejudice the value of the trial result. Therefore we used blinded OPEN trial participant outcome data to calculate the SD of the primary outcome measure from first 69 OPEN participants who have submitted at least one post-operative measure (220 measurements in total). The empirical SD was 0.15, considerably smaller than the assumed value of 0.33 used in our initial calculations. As this was based in immature data we expect variability to increase over time. To allow for this we have assumed a conservative SD of 0.21, giving a revised sample size of 170 randomised men with complete follow- up inflated to 200 in total to allow for loss to follow-up (a conservative 15% rate of attrition).

The trial is also powered to determine whether the use of urethroplasty will result in a 30% (from 50% to 20%) reduction in need for further intervention at two years compared to urethrotomy as a secondary outcome. To detect this difference using the binomial test of proportions with 90% power at the 5% significance level would require 52 men to complete the study in each arm, giving a total of 104 men.

3.3. Levels of confidence

Statistical significance will be at the 2-sided 5% level with corresponding 95% confidence intervals (CI) derived.

1.1. Interim analysis

No interim analysis is planned for this study.

1.2. Time points

				Months after surgery				Months after randomisation			
	Baseline	Prior to surgery	1 week after catheter removal	3	6	9	12	18	24	Prior to/after any surgical re-intervention	End of study (same calendar time for every participant)
ICIQ-Male SF	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
IIEF male sexual QoL	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
EQ-5D	✓	✓	✓	✓	✓	✓	✓				✓
Urinary flow rate	✓			✓			✓		✓		
Rate of recurrence								✓	✓		✓
Need for further intervention								✓	✓		✓

2. Analysis

2.1. Statistical methods

2.1.1. Analysis method

All the main analyses will be based on the intention-to-treat principle (i.e. analyse as randomised), although additional analysis groups such as per-protocol will be considered. Baseline and follow-up data will be summarised using mean (standard deviation) or median (interquartile range) where appropriate for continuous variables. Discrete variables will be summarised with numbers and percentages. Treatment effects will be presented with 95% confidence intervals.

2.1.2. Primary outcome

The primary outcome measure, area under curve (AUC) for ICIQ-Male Short Form (ICIQ-Male SF) questionnaire over 24 months following randomisation will be analysed using linear regression adjusted for minimisation covariates. Measurements are taken at baseline; prior to surgery; 1 week after catheter removal; 3, 6, 9 and 12 months after surgery; 18 and 24 months after randomisation. We also have additional measures taken pre and post at re-intervention, 24 months post-surgery and a final measure at the end of study. The AUC will be constructed using the trapezoidal rule, which assumes a constant increment (or decrement) in score between two points where outcome is measured.

The primary outcome measure will be analysed using linear regression with adjustment for the minimisation variables [site of recruitment and time since last procedure (<12 months or ≥12 months)].

Our primary analysis will be on observed data, to meet inclusion in this analysis participants must have at least three measures of the ICIQ-Male SF, one at baseline, one "earlier" and one "later". In more detail, we require:

- A baseline measure of outcome, if this is missing the centre-mean baseline outcome measure will be imputed to calculate the AUC.
- An early measurement, i.e. a measurement taken at least one of 3, 6, 9 or 12 months post-surgery.
- A later measurement, i.e. a measurement taken at least one of 18 or 24 months post-randomisation, 18 or 24 months post-surgery.

Sensitivity analyses will be conducted to assess the robustness of the treatment effect estimate to this approach by relaxing and tightening the minimum number of measures needed. Where available all observed data will be used, however it is likely that missing data will be present at various time points. The assumptions of the primary analysis and proposed sensitivity strategies are outlined for each group of measurement time points.

Baseline and prior to surgery: We anticipate minimal missing data at these time points. If either measure is missing we will assume a constant score between these two time points and impute one with the other. We will be able to assess empirically on observed data if this is a reasonable assumption. If not, sensitivity analyses will use imputation (under MAR assumption) and pattern mixture approach (MNAR). If both are missing we will impute the centre mean for each time point to allow calculation of the AUC.

1 week after catheter removal: If this measure is missing we will not impute a value, the AUC calculation will be between prior to surgery and the first early measure.

3, 6, 9 and 12 months post-surgery: We will use all available information but require only one of these time points to calculate the AUC. The AUC calculation will use the notional time (in weeks) between the last available time point prior to 3 months and the first of these time points. If one or two time points are missing, we will not impute a value for those missing time points but assume constant increment (or decrement) in score between points where outcome is measured. Sensitivity analyses for missing data at these time points are covered in the general sensitivity analysis approach below.

18 and 24 months post-randomisation: We will use all available information but require only one of these time points to calculate the AUC. If the 18 month time point is missing but the 24 month time points is measured, we will not impute a value, the AUC calculation will use the notional time (in weeks) between the last available measurement prior to 18 months 24 months and assume constant increment (or decrement) in score between these two time points. If the 18 month time point is measured but the 24 month time point is missing we will carry the 18 month measurement forward to 24 months to allow the calculation of the AUC. If we have a 24 month post-surgery or end of study measurement that is closer to 24 months (i.e. less than 6 months difference) we will use that measure rather than the 18 month measure.

Pre- and post-intervention: Where participants receive re-intervention and have observed outcome measures we will incorporate these into the relevant time section of the AUC, i.e. inserting the extra observations between notional time point measures. Where re-intervention clashes with an expected outcome measurement the re-intervention reported measurements will be used. If re-intervention takes place but the outcome measures are missing we will use index intervention outcome data for that participant. We can assess this assumption empirically with observed data.

Missing follow-up data will be estimated in sensitivity analysis using multiple imputation models for participants that meet the minimal follow-up requirements but have missing time points. We will explore differences between responders and non-responders to inform our missing data model. We will calculate an AUC for each imputation and combine using Rubin's rules under a MAR random assumption. We will also explore, using pattern mixture models to impute a range of values (to be estimated from observed data) different MNAR scenarios. Participants that don't meet the primary analysis criteria will be included in an additional sensitivity analysis.

Measures of the primary outcome collected at 24-months post-intervention and at the end of the study will also be included, when applicable, as a sensitivity analysis of the calculation of the AUC.

2.1.3. Secondary outcomes

The following secondary outcomes will be recorded at baseline, immediately prior to surgery, 1 week after catheter removal and 3, 6, 9, 12, 18 and 24months after surgery and 18 and 24 months after randomisation, at end of study and prior and subsequent to any surgical re-intervention:

1. Difference in **condition-specific quality of life** trajectory measured by the AUC for the single item ICIQ-MaleSF QoL score.
2. Difference in **global sexual functioning** trajectory measured by the AUC for the single item male sexual QoL score.
3. Difference in **generic quality of life** trajectory measured by the AUC for the EQ-5D (5L version) total score based upon responses to 5 dimension items and using UK population valuations (0 death to 1 full health) and visual analogue scale (VAS) score (0 worse possible health state – 100 best possible health state). Not recorded prior or subsequent to any surgical re-interventions.

Other secondary outcomes:

1. Difference in rate of improvement of **urinary flow rate** measured as part of routine care at baseline, 3 and 24 months with an increase in $Q_{max} \geq 10$ ml/s from baseline taken to signify a successful outcome[3].
2. Difference in rate of **rate of recurrence and need for further intervention** recorded from the clinical record for those returning to the care of their original specialist with recurrent stricture, by patient questionnaire for participants seeking care elsewhere, and checked by the local trial research staff at the final 24 month assessment. For participants in whom the clinical record documents stricture recurrence the relevant clinical information will be sent in anonymised form as a case vignette to an expert panel of urology clinicians independent of the trial to determine whether or not there is a majority opinion that clinical recurrence of the stricture has been confirmed.

Secondary outcomes will be analysed using generalised linear models appropriate for the distribution of the outcome with adjustment for minimisation and baseline variables as appropriate. Further analysis will explore the impact of variations in treatment delivered; such as use of anastomotic urethroplasty and use of intermittent self-dilatation after urethrotomy.

From the feasibility phase estimates of recruitment rates and potential participant availability will be reported, together with appropriate confidence intervals. There are no planned interim outcome analyses; all analyses will occur following completion of trial follow up. Interim analyses will be performed if requested by the Data Monitoring and Ethics Committee (DMC).

2.1.4. Subgroup analysis

Subgroup analyses will explore the possible modification of treatment effect by clinically important factors; time since last procedure (<12 months or ≥ 12 months) as a global measure of stricture severity, age, stricture location, and length. This will be done by including treatment-by-factor interactions in the model and they will be classified as exploratory analyses.

3. Safety data

An adverse event (AE) is defined as any untoward medical occurrence in a subject to whom a study intervention or procedure has been administered, including occurrences which are not necessarily caused by or related to that intervention.

For purposes of this protocol:

- All adverse events will be recorded at time of primary or re-intervention surgery, 3 months, 12 months and 24 months and categorised as to expectedness, relatedness and severity.
- Any serious adverse events will be recorded throughout the duration of the trial
- 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.
- Serious adverse events exclude stricture (symptom or urine flow) recurrence which is already documented and monitored within study.

Please refer to the protocol for more information about AE and reporting AE.

4. Statistical software

The most recent version of Stata at the time of analysis will be used.

5. Dummy tables

5.1. Baseline patient characteristics

	Allocated Urethrotomy	Allocated Urethroplasty
Age – Mean (sd)		
Time since last procedure – n (%) < 12 months >= 12 months		
Urine flow rate (ml/s) – mean (sd)		
ICIQ-Male SF symptom scores – median (p25, p75) (0-4, where 0 is no symptoms and 4 is continuous symptom) Delay before start to urinate Strength of urinary stream Strain before urinating Stop & start whilst urinating Feeling bladder has not emptied properly after urinating Frequency of wetting of pants after finishing urinating		
ICIQ-Male SF total score 0 - 24 – mean (sd)		
ICIQ-Male SF QoL score (0-3) – median (p25, p75)		
IIEF male sexual QoL score (1-5) – mean (sd)		
Stream picture score (1-4) – median (p25, p75)		
EQ5D – n (%) respondents Mobility: No problems / Some problems Self-care: No problems / Some problems Usual activities: No problems / Some problems Pain/discomfort: No problems / Some problems Anxiety/depression: No problems / Some problems		
EQ5D total score based on above 5 dimensions – mean (sd)		
EQ5D VAS – mean (sd)		

5.2. Trial intervention

Variable	Units	Allocated Urethrotomy	Allocated Urethroplasty	Difference CI, P value
Interval between randomisation and intervention	Days (mean SD)			
No Intervention	Number participants (%)			
Received urethrotomy	Number participants (%)			
Received urethroplasty	Number participants (%)			
Hospital stay for trial intervention	Hours (mean SD)			
Duration of procedure	Minutes (mean SD)			
Antibiotic prophylaxis used	Number participants (%)			
Grade of Surgeon	Number participants (%)			
Consultant				
Career (Staff) Grade				
Trainee				
Length of stricture (operative estimate)	mm (mean SD)			
Duration catheterisation	Hours (mean SD)			

5.3. Outcomes

Variable	Units	Allocated Urethrotomy		Allocated Urethroplasty		Difference CI, P value
AUC symptom score over 24 months						
Participant score of urine flow at 24 months	Number (%) at each value					
1						
2						
3						
4						
Condition bother score at 24 months	Number (%) at each value					
Not at all						
A little						
Somewhat						
A lot						
AUC of Condition bother score over 24 months	Mean (SD)					
AUC of EQ5D over 24 months	Mean (SD)					
AUC of global sexual function score over 24 months	Mean (SD)					
Re-intervention up to 24 months	Number participants (%)					
Time to re-intervention	Weeks (mean SD)					
Re-interventions at 24 months	Number of re-interventions (mean SD)					
Qmax at 24 months	mL/sec (mean SD)					
≥ 10 mL/s improvement in Qmax at 3 months	Number participants (%)					
≥ 10 mL/s improvement in Qmax at 24 months	Number participants (%)					

Condition bother score at 24 months						
Recurrence of stricture up to 24 months	Number participants (%)					

5.4. Adverse effects

Variable	Units	Allocated Urethrotomy	Allocated Urethroplasty	Difference CI , p-value
Complications during operation	Number of participants (%)			
Post-Operative complications in hospital	Clavien Grade number participants (%)			
Clavien 1 or 2				
Clavien 3 or 4				
Death				
Urogenital infection up to 3 months after discharge	number participants (%)			
Other AEs up to 3 months	number participants (%)			
Satisfaction with sex life at 24 months	number participants (%)			
Very satisfied				
Moderately satisfied				
About equally				
Moderately dissatisfied				
Very dissatisfied				
Deaths up to 24 months	number participants (%)			