NeuroSAFE PROOF: Statistical Analysis Plan

A single blinded, IDEAL stage 3, multi-centre, randomised controlled trial to assess NeuroSAFE Robotic assisted radical prostatectomy (RARP) vs standard Robotic assisted radical prostatectomy (RARP) in men with prostate cancer

Version 2.1 18Sep24

Preface

This statistical analysis plan has been prepared by:

Dr Shengning Pan and Dr Baptiste Leurent

Department of Statistical Science
University College London
Gower St.
London WC1E 6BT

E-mail: shengning.pan.18@ucl.ac.uk

With support from:
Prof Greg L. Shaw, Mr Nick Roberts, Dr Ricardo Almeida Magana and Dr Eoin Dinneen

Approval

Senior statistician

Name	Dr. Baptiste Leurent
Position	Associate Professor in Medical Statistics, University College London
Signature	
Protocol version	2.1 18Sep24
Date	18 /09/24

Chief Investigator

Name	Prof. Greg Shaw
Position	Professor of Urology, University College London Lead for Robotic Urology, University College London Hospital
Signature	
Protocol version	2.1 18Sep24
Date	18/09/24

Contents

1 Trial summary

NeuroSAFE is a single blinded, IDEAL stage 3, multi-centre, randomised controlled trial to assess NeuroSAFE Robotic assisted radical prostatectomy (RARP) vs standard RARP in patients with prostate cancer. Patients will be randomised 1:1 to NeuroSAFE RARP (intervention group) vs Standard RARP (control group). This trial aligns with an IDEAL framework Stage 3 Assessment for the evaluation of complex surgical interventions.

1.1 Summary table

Scientific Title	A single blinded, IDEAL stage 3, multi-centre, randomized controlled trial to assess NeuroSAFE Robotic assisted radical prostatectomy (RARP) vs standard RARPs in men with prostate cancer
Short Title	NeuroSAFE at prostatectomy to optimize oncological and functional outcome. (NeuroSAFE PROOF)
Trial registration	NCT03317990
Aim	This is a fully powered RCT to assess the differences in functional and oncological outcomes between RARP using the NeuroSAFE intraoperative frozen section technique to navigate safe nerve sparing and standard UK practice RARPs
Countries of Recruitment	UK
Health Condition(s) or Problem(s) Studied	Robotic assisted radical prostatectomy for prostate cancer
Intervention	Control group - standard RARP: Nerve-sparing strategy during RARP guided by routinely available pre-operative and intra-operative patient/disease information; mpMRI, biopsy tumour histology, DRE. Intervention group – NeuroSAFE RARP: Nerve-sparing strategy during RARP guided by the NeuroSAFE technique during RARP.
Inclusion criteria	- Men opting to undergo RARP for organ confined prostate cancer (including radiological t3a). - Potent men (IIEF score of 22-25 on the first 5 questions of the IIEF-15 without any erectile function medical assistance) - Men who are continent of urine (no self-reported urinary incontinence) - Has given written informed consent - Ability to read English sufficiently to answer questionnaires and understand PIS

Exclusion criteria	- Unable to undergo RARP - Known overactive bladder - Any previous treatment for prostate cancer - Previous/current hormone treatment for prostate cancer - Nerve sparing deemed futile due to locally advanced disease by surgeon and radiologist
Number of sites	5
Study setting	Secondary care
Date of First Enrolment	10.10.18
Target Sample Size	Patient recruitment target in protocol: 404 patients Loss of follow-up: 10%
Primary Outcome	- IIEF-5 score at 12 months
Secondary Outcomes	- ICIQ at 3 months - ICIQ at 6 months - IIEF-6 at 12 months
Descriptive outcomes	- Time to erective function recovery - Time to continence recovery - Positive surgical margins - Oncological outcomes at 12 months

1.2 Objectives

1.2.1 Primary objective

To assess the difference in erectile function between patients undergoing standard RARP (control group) and NeuroSAFE RARP (intervention group) at 12 months following treatment.

1.2.2 Secondary objectives

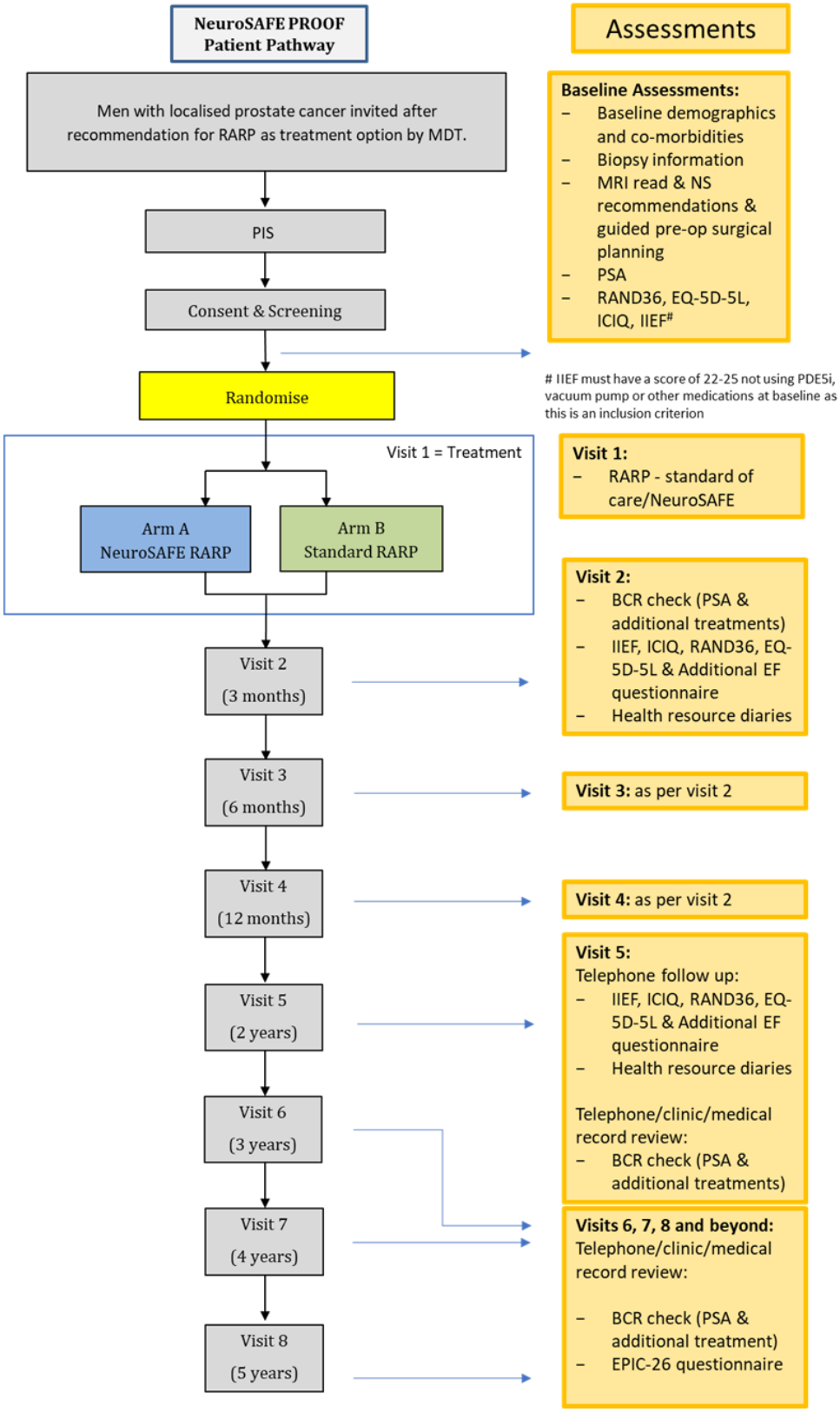
1. To assess the differences in urinary continence between patients undergoing standard RARP and NeuroSAFE RARP in the early period following treatment (3 and 6 months).

2. To report the oncological outcomes between patients undergoing standard RARP vs. NeuroSAFE RARP at 12 months.

3. To evaluate differences in overall quality of life outcomes between patients undergoing standard RARP vs NeuroSAFE RARP. (Not covered in this SAP)

4. Economic analysis to assess health resource use between standard RARP vs NeuroSAFE RARP at 12 months and up to 2 years post treatment. (Not covered in this SAP).

1.3 Trial diagram



1.4 Sample size

The sample size was based on detecting a 14% difference in the proportion of patients with an IIEF-5 score of 21 or above between the intervention (NeuroSAFE RARP) and control (standard RARP) groups (42% vs. 28%). 364 evaluable participants are required to have 80% power to detect such difference, with a two-sided alpha of 5%. Allowing for 10% drop out, the trial aimed to recruit 404 participants (202 per arm).

1.5 Randomisation

Block randomisation, stratified by site with 1:1 allocation. Randomisation was performed using the online based service Sealed Envelope™.

2 SAP introduction

2.1 Purpose of statistical analysis plan

This Statistical Analysis Plan (SAP) reports the details of the main effectiveness analyses of the NeuroSAFE randomised controlled trial.

Its purpose is to reduce bias in the analysis and presentation of study results by pre-specification of the analyses conducted and the statistical methods used, with an appropriate level of detail.

This SAP does not prevent adjustments to the statistical methodology during data analysis, provided that these adjustments are limited and justified. It also does not prevent conducting further secondary analyses that appear relevant after conducting the pre-specified analyses, these will also be limited, justified, and clearly identified.

This SAP covers only the main effectiveness (clinical) analysis of the trial up to the visit 4 follow-up (12 months). Longer term outcomes, quality of life and health economic evaluation will be covered in separate analysis plans.

2.2 Blinding status of personnel involved in SAP

This report has been written by Shengning Pan (trial statistician) and Baptiste Leurent (senior statistician) with support from Greg L. Shaw (Chief Investigator), Nick Roberts (Trial manager), Ricardo Almeida Magana (Clinical research Fellow) and Eoin Dinneen (Clinical research Fellow).

SP and BL are blinded to allocation status and will remain until publication of the SAP on a public repository and the final analysis is ready to be conducted.
NR was unblinded as part of his trial manager role.
GS, RAM, and ED were in principle blinded to the allocation status. They were however aware of the allocation for the individual participants they operated. They could also potentially access data which could reveal the allocation of individual participants, for example by looking at the operations information in the study dataset. They however did not have access to the allocation variable, nor conducted any comparison between arms. One exception is ED who had access to unblinded data for the 120 first participants as part of his PhD research. His thesis was shared with GS but no other members of the teams.

SP and BL were responsible for writing this SAP, based on the study protocol and an earlier version of the SAP, and in collaboration with NR, GS, RAM and ED who had an advisory role.

2.3 Unblinding and data analysis

This SAP will be approved by BL and GS and published on a publicly available repository (such as figshare.com orClinicalTrials.gov registry) before unblinding and final analysis.

Unblinding of data and analyses will commence after the last patient has completed the 12-month follow-up, all relevant data has been entered, data checks have been performed and the final analysis plan has been confirmed and approved. Majority of the analytical programs will be prepared prior to unblinding, including at least the primary outcome analysis.

A copy of the dataset at the time of the unblinding will be archived.

2.4 Data checking

Check will be conducted before the analysis to identify missing or implausible data. These checks include:

- Missing data
- Data outside expected range
- Inconsistency between dates, e.g. baseline data recorded after visit2.

Where abnormalities are identified, the data will be double-checked with the Clinical Trial Manager, who will liaise with the study sites who will enter the data directly to the study database

3 Data collection and outcomes definition

3.1 Schedule of assessments

Visit	Screening/ Baseline	Random-isation	Treatment	Visit 2 3 months	Visit 3 6 months	Visit 4 12 months
Informed consent	x					
Randomisation		x				
Fitness for surgery assessment	x					
PSA	x			x	x	x
Standard care referral pathway to regional uro-oncology centre	x					
MRI guided pre-op surgical planning			x			
RARP (standard/control or NeuroSAFE/intervention)			x			
Adverse events			x	x		
EQ-5D-5L – (in clinic)	x			x	x	x
ICIQ – (in clinic)	x			x	x	x
IIEF – (in clinic)	x			x	x	x
RAND36 – (in clinic)	x			x	x	x
Additional Erectile function questionnaire				x	x	x
Adjuvant Treatments Assessment				x	x	x
Health Resource Diaries data collection				x	x	x

The study collected additional data through electronic PROMs, paper PROMs or telephone consultation at 2 years (RAND36, IIEF, EQ-5D-5L, ICIQ and Additional EF questionnaire), and at 3, 4 and 5 years via additional medical record review (EPIC-26, PSA & Adjuvant therapies) and telephone or clinic (BCR rates). These longer-term data are not covered in this SAP.

3.2 Outcomes list

Primary outcome:

- IIEF-5 (the simplified International Index of Erectile Function) at 12 months.

Secondary outcomes:

- ICIQ (International Consultation on Incontinence Questionnaire) at 3 months *
- ICIQ at 6 months
- IIEF-6 at 12 months

* Main secondary outcome of interest is ICIQ at 3 months.

Descriptive outcomes:

- Erectile function recovery at 3, 6 and 12 months (binary and categorical)
- Time to erectile function recovery
- Continence recovery at 3, 6 and 12 months (binary)
- Time to urinary continence
- Oncological outcome at 12 months
- Positive Surgical Margins

Additional descriptive data will also include baseline characteristics, surgery and histopathology information, and serious adverse events, as described below.

3.3 Time points and dates

Data were aimed to be collected at baseline, 3 months after surgery (visit2), 6 months after surgery (visit3) and 12 months after surgery (visit4).

Outcomes were not always collected at the exact time point, particularly during the COVID-19 pandemic, and we allow some flexibility around the due date to determine if data could contribute to the analysis. Assessments were considered valid if they were collected within the following time windows:

- Visit 2 (3months): Recorded as visit 2, and conducted between 1 to 6 months after surgery
- Visit 3 (6 months): Recorded as visit 3, and conducted between 4 to 12 months after surgery
- Visit 4 (12-months): Recorded as visit 4, and conducted between 9 to 18 months after surgery

Date of PSA measurement available could vary more, and no specific window was applied. All available PSA data were used to determine the oncological outcome to maximally capture any cases of oncological failure.

Handling of missing visit dates:

If the questionnaire date was missing (for IIEF and ICIQ), we will look at the corresponding date for the other questionnaires. If the date of the other questionnaires were also missing, we will impute the date assuming the observation was recorded at the expected time point from surgery.

If the PSA date was missing, we will assume the observation was recorded at the expected time point from surgery.

Treatment date was recorded as day, month, year. If the day was missing, we will assume it was done on the 15th of the month. If the month or full date was missing, we will assume treatment has started on the expected visit date, checking this is consistent with the PSA and other information available.

3.4 Primary outcome definition

The primary outcome is the simplified International Index of Erectile Function (IIEF-5, Rosen 1999), at 12 months. The IIEF-5 is also known as the Sexual Health Inventory for Men (SHIM).

The questionnaire used in the study was based on the 15-item British Association of Urological Surgeons (BAUS) version of the IIEF (see Appendix 6.1).

The IIEF-5 score is obtained by summing the scores to the questions 2, 4, 5, 7 and 15, as described in Appendix 6.1.

The IIEF-5 score ranges from 1 to 25, with a higher score indicating better erectile function.

Handling of partially complete IIEF-5
For participants who answered at least half of the questions, we will replace any of the missing answers by the lowest score (0 for questions 2, 4, 5, and 7, and 1 for question 15), thus returning a conservative total score. For participants with 3 or more missing responses, we will consider the IIEF-5 score as missing.

This IIEF-5 definition will be also applied at baseline, visit 2 and visit 3.

3.5 Secondary outcomes definition

3.5.1 ICIQ

Incontinence was assessed using the International Consultation on Incontinence Questionnaire (ICIQ) (Avery, 2004).

The ICIQ score is obtained by summing the questions 3, 4 and 5 of the ICIQ questionnaire (see appendix 6.3).
The ICIQ score ranges from 0 to 21, with higher scores representing worse incontinence.

Handling of partially complete ICIQ
Any missing question 3 or 4 will not be imputed. For participants with missing question 5, if their answer was 0 (“Never leaks urine”) at question 3 and 4, we assumed the answer at question 5 was 0 (“Not at all”).
Other partially complete response will not be imputed, and total ICIQ score considered as missing.

This ICIQ score definition will be applied at baseline, visit 2, visit 3 and visit 4.

3.5.2 IIEF-6

One of the secondary outcomes will be IIEF-6 at 12 months.

IIEF-6 score (Cappelleri J. C., 1999) will be derived by adding the scores to the question 1, 2, 3, 4 ,5 and 15 of the IIEF, as described in Appendix 6.2.

Handling of partially complete IIEF-6
The missing data will be handled similarly as the IIEF-5, replacing missing answers by the lower score for those completing at least 3 of the 6 IIEF-6 questions.

This IIEF-6 definition will be also applied at baseline, visit 2 and visit 3.

3.6 Descriptive outcomes definition

3.6.1 Erectile function recovery

Erectile function recovery (binary outcome) will be defined at each time-point, based on whether the IIEF-5 score is larger or equal to 15.

We will also report at each time-point the IIEF-5 as a categorical variable, as suggested in Rosen 1999: Severe (1-7), moderate (8-11), mild-moderate (12-16), mild (17-21), and no ED (22-25).

3.6.2 Time to erectile function recovery

This will be defined as the time from surgery to the time of first reaching an IIEF-5 score larger or equal to 15.

3.6.3 Urinary continence recovery

Urinary continence recovery (binary outcome) will be defined at each time-point, based on whether the ICIQ score is less or equal to 5.

3.6.4 Time to urinary continence recovery

This will be defined as the time from surgery to the time of first reaching a ICIQ score of 5 or below.

3.6.5 Oncological outcome

Oncological outcome after 12 months (categorical outcome) will be classified into four categories according to the following definition:

1. <u>PSA persistence</u>
   PSA≥0.2 at visit 2

2. <u>Biochemical recurrence</u>
   PSA<0.2 at visit 2
   AND
   (PSA≥0.2 at visit 3 **OR** PSA≥0.2 at visit 4)

3. <u>Early salvage treatment</u>
   PSA<0.2 at each of visit 2, visit 3, and visit 4

AND
Received adjuvant treatment* in the first year

4. Other (No recurrence or treatment)
   PSA<0.2 at each of visit 2, visit 3, and visit 4
   AND
   No adjuvant treatment* in the first year

*Adjuvant treatment is defined as any adjuvant cancer treatment during the first year after surgery, as recorded on the study questionnaire at visit2, visit3 and visit4, under the question "Did the subject have any or have any adjuvant therapy planned?". Common adjuvant treatment could include Hormonal Treatment or Radiotherapy.

Handling of partially missing information for oncological outcome
Participants with no adjuvant treatment reported will be considered as not receiving adjuvant treatment. Missing treatment date will be handled as discussed above (see section 3.3).
Participants with undefined oncological status due to missing PSA values will be reviewed case by case. If a participant has a PSA below 0.2 at a visit and has not started treatment since, we will assume previous PSA values were below 0.2. We will also use the later available PSA values (visit 5, part of the longer-term follow-up) to apply this rule.
For patients who died during the 12 months follow-up, we will report their oncological status before death (if available) but indicating that these patients have died, with the reason of death.

3.6.6 Positive Surgical Margins (PSM)

Participants surgical margins will be classified into three categories according to the following definition:

1. Negative surgical margins
   No positive surgical margins

2. Small single positive surgical margin
   Positive surgical margin in a single location
   AND
   Length of the positive margin is less than 3mm

3. Large or multifocal surgical margin
   Positive surgical margin in at least two locations
   OR
   Length of any positive margins at least 3mm

Handling of missing PSM information
Any participant who was operated but do not have PSM data, or sufficient data to be classified into one of the three categories will be considered as missing.

3.7 Baseline covariates definition

- Age (continuous)
  Age as recorded at the time of consent, in years

- Site (categorical)
  Five categories: UCLH, Sheffield, Bristol, Glasgow and Nottingham.

Baseline IIEF-5, ICIQ and IIEF-6 are defined above.

3.8 Subgroup definition

A subgroup analysis will be conducted according to the baseline pre-operative magnetic resonance imaging (MRI) recommendation for nerve sparing.

This will be defined as follow:
Bilateral nerve sparing recommended (BNSR) will be defined as having pre-operative nerve spare recommended on both the right AND on the left side by the MRI planning meeting.

Participants without BNSR (the main subgroup of interest) will be defined as those for whom nerve sparing was not recommended, or digital rectal examination (DRE) was recommended to guide the degree of nerve sparing, on either (or both) of the right and left sides. This would therefore include any of the following recommendations:

- (NS not recommended or DRE recommended) on the right side
- (NS not recommended or DRE recommended) on the left side
- (NS not recommended or DRE recommended) on both sides

Handling of partially missing information for BNSR
Participants with at least one of the side missing their NS recommendation will be considered as missing for BNSR.

3.9 Other descriptive variables

We list here briefly other relevant variables; the exact list and definition may be adapted at time of publication.

3.9.1 Baseline demographic characteristics

- Co-morbidities
  Proportion of patients with any comorbidity (binary)
  Proportion of patients with the four most frequent comorbidities (binaries)
- Ethnicity (categorical)
  The ethnicity will be reclassified as White, Black, Asian and Other/Mixed.
- Body Mass Index (BMI, continuous)

Defined as weight(kg)/height(m)^2

- Smoking status (categorical)

  The smoking status classified as current smoker, ex-smoker, non-smoker and missing/unknow.

3.9.2 Baseline oncological characteristics and MRI

- Prostate volume (continuous)
- Baseline PSA (continuous)
- Histological classification
  - Gleason Grade Group (also known as International Society of Uropathologists (ISUP) grade group 1 to 5 (see table Appendix 7.4)
- Total cores (continuous)
- Positive cores

  The percentage of positive cores is defined as percentage of positive cores = number of samples that were positive *100 / total number of samples taken during biopsy
- Maximum core length (continuous)
- Clinical T stage

  Categorical: 1,2, or 3
- Cambridge prognostic group

  Categorical: 1, 2, 3, 4 , 5.

  See Appendix for detailed definitions.

3.9.3 Surgery procedure

- Overall surgical time (continuous)

  Defined as the time from knife to skin to skin closed
- NeuroSAFE time (continuous)

  Defined as the time between specimen removed for painting and decision phoned to surgeon)
- Lymph node dissection (Yes, No, Unknown)
- Nerve sparing outcomes
  1. By patient level (Bilateral vs Unilateral vs None)
  2. Quality of nerve spare on right (Grade 0, 1, 2, 3)
  3. Quality of nerve spare on left (Grade 0, 1, 2, 3)
- For NeuroSAFE arm only
  1. Negative NeuroSAFE

     Defined as frozen sections R1 - R10 and L1 - L10 ALL equal to ‘clear’ or ‘narrowly clear’
  2. Positive NeuroSAFE

     Defined as ANY positive result in frozen sections (R1 - R10 and L1 - L10).
  3. Nerve bundles resected

     Defined as patient taking right OR left bundle action

3.9.4 Pathology results

- Prostate weight (continuous)
- Histological classification

    - Gleason Grade Group (also known as International Society of Uropathologists (ISUP) grade group 1 to 5 (see table Appendix 7.4)
- pT stage (Categorical)
    - Defined as pT2, pT3a, pT3b or pT4
- pN stage (Categorical)
    - Defined as pNx, pN0, or pN1
- Positive margins
    - See above for definitions
- Concordance to frozen section (NeuroSAFE arm only)
    - Frozen section concordant with final analysis of postero-lateral margin
- Presence of tumour in secondary resection (binary)
    - Cancer noted in the right or left secondary resection

4 Statistical analysis

4.1.1 Analysis populations

Intention-to-treat population (ITT) population: all participants who were enrolled in the trial and randomised.
Modified intention-to-treat (mITT) population: all participants who were randomised and received treatment (standard RARP or NeuroSAFE RARP)

Unless specified otherwise, analyses will be on mITT population

4.1.2 Handling of missing data

Frequency of missing data will be reported descriptively, as explained in section 4.6. We will explore baseline characteristics associated with missingness, using logistic regression on missing IIEF-5 at 12 months as outcome.

Handling of partially incomplete variables
In section 3.3 (Time points and dates), 3.4 and 3.5 (Outcomes), we explain how we will handle some of the partially missing observations. For example, if participants have responded at least 3 or the 5 IIEF-5 questions, we will replace the missing questions by 0 to obtain a conservative IIEF-5 score. These ad-hoc imputations when constructing the variables are all described in the outcome definition, in section 3

Handling of missing baseline covariates
For the outcomes adjusted analysis, if some of the baseline adjustment covariates (site, age, baseline IIEF-5, ICIQ or IIEF-6) are missing we will impute them using mean or mode imputation (White, 2005).

Missing outcomes
The main analysis (of the primary and secondary outcomes) will be based on all available data, under the missing-at-random assumption, i.e. that the probability of being missing is independent of the unobserved outcome, conditionally on the observed data in the model (arm and baseline covariates). Sensitivity analysis under different assumptions will be conducted, as described in 4.11.

4.1.3 Hypothesis testing

All statistical tests will be two-sided and considered significant at the 5% significance level, unless otherwise specified.
P-values reported will be for the null hypothesis of no difference between the two randomisation groups (NeuroSAFE RARP vs. standard RARP)
There is a limited number of tests conducted (one primary and three secondary outcomes, tested overall and within a subgroup), and no formal correction for multiple testing will be applied. The number of comparisons performed will be taken into account informally when interpreting the results.

Confidence interval reported will be at the 95% level, two-sided.

4.2 Participants flowchart

A flowchart will report the number of patients at each stage of the trial following the CONSORT guidelines.
The flow chart will report by arm: the number of participants randomised, receiving the surgery, completing visit 2, 3 and 4, and the reasons for early withdrawal from the trial.

4.3 Baseline characteristics

The baseline demographic and clinical characteristics (see list in section 3.9.1) will be summarised descriptively, overall and by randomisation arm, on the mITT sample. Mean, median, standard deviation, interquartile range, range (for continuous variables), or frequency and percentage (for categorical variables), will be reported as appropriate.

Baseline characteristics of participants who did not complete surgery will also be reported as additional information.

4.4 Surgery and pathology description

The surgery details and pathology results (see list of variables in sections 3.9.3 and 3.9.4) will be summarised descriptively by randomisation arm, on the mITT sample.

4.5 Adverse events

The number and proportion of participants who experienced any adverse events and any serious adverse events within 90 days from surgery will be reported by arm (frequency and proportion).
A table will describe the type of SAE experienced. We will also report how many were considered possibly related to the NeuroSAFE intervention.

4.6 Descriptive statistics

We will provide summary statistics for all study variables, overall and by arm, at the different time points. Summary statistics for continuous variables will include mean, median, standard deviation range and/or interquartile range as appropriate. Categorical variables will be described using frequencies and percentages. Main variables (outcomes listed in section 3.2) will be described in the main report, while additional descriptive statistics will be reported in an appendix. Graphical representation may be considered.
The appendix will also report the frequency of missing data for all variables.

4.7 Primary outcome analysis

The primary estimand of interest will be the difference in mean IIEF-5 score at 12 months between participants allocated to the NeuroSAFE and those allocated to the standard treatment group among those who received the surgery.

The primary outcome (the IIEF-5 score at 12 month) will be analysed using a normal linear regression model. The model will include randomisation group, site (the stratification variable), age, and IIEF-5 at baseline. Continuous covariates (age and IIEF-5) will be included as a linear term, while site will be included as categorical. The estimated coefficient for the adjusted mean difference between randomisation groups will be presented alongside its 95% confidence interval and two-sided p-value

The primary analysis will include everyone in the mITT population with IIEF-5 available at 12 months.

4.8 Secondary outcomes analysis

The secondary outcomes (ICIQ at 3 months, ICIQ at 6 months, IIEF-6 at 12 months) will be analysed using linear regression analogous to the primary outcome analysis. The models will include randomisation group and site, age and corresponding outcome (ICIQ or IIEF-6) score at baseline. It will be based on the mITT patients with relevant outcome data available.

4.9 Descriptive outcomes analysis

4.9.1 Positive Surgical Margins

The PSM will be reported descriptively, reporting the number and percentage of participants in each of the PSM category, by arm.

4.9.2 Time to erectile function recovery

Descriptive analysis of time to erectile function recovery (IIEF-5 scores larger or equal to 15) will be performed displaying the Kaplan Meier survival curve by arm.

IIEF5 data included for this analysis will be those from visit 2 to visit 4, as described in section 3.3. Follow-up time will start on day of surgery and until visit 4. Patients will be assumed as not recovered at entry. Patients will be censored when first reaching IIEF-5≥15 or at date of last available IIEF-5.

Kaplan Meier curves will be shown until 15 months.

4.9.3 Time to urinary continence

Descriptive analysis of time to urinary continence (ICIQ scores less or equal to 5) will be performed displaying the Kaplan Meier survival curve by arm.

ICIQ data included for this analysis will be those from visit 2 to visit 4, as described in section 3.3. Follow-up time will start on day of surgery and until visit 4. Patients will be assumed as incontinent at entry. Patients will be censored when first reaching ICIQ≤5 or at date of last available ICIQ.

Kaplan Meier curves will be shown until 15 months.

4.9.4 Oncological outcome

We will report descriptively the oncological outcome during the first year, presenting the frequency and percentage in each category, by arm.

4.10 Subgroup analysis

4.10.1 Rationale for the subgroup analysis

An important subgroup for this study is men who did not receive a pre-operative radiologist BNSR. We hypothesised that the effect of NeuroSAFE will be more beneficial in patients who did not receive a recommendation for bilateral nerve sparing. This is because our earlier analysis (Dinneen E. a.-H.-P., 2022) have demonstrated that when a patient is recommended a bilateral nerve spare based on the tumour appearing to be away from the capsule of the prostate, this is almost always correct, such that these patients to do need to have NeuroSAFE in order for the objective of maximal safe nerve sparing to be achieved. The non-BNSR subgroup therefore consists of those patients in whom NeuroSAFE offers a potential benefit.

4.10.2 Interaction test

We will assess whether the effect of NeuroSAFE on IIEF-5 at 12 months and ICIQ at 3 months differs according to pre-operative bilateral nerve sparing recommendation (BNSR), as defined in 3.9.2.

This will be assessed by testing for an interaction between BNSR and treatment arm in the primary and secondary outcome models (linear regression adjusted by site, age and baseline IIEF-5 or ICIQ).

Interaction test p-values will be interpreted in consideration of the low power of the interaction test, the number of subgroups tested, and the biological plausibility of the interaction.

4.10.3 Subgroup analysis within the non-BNSR participants

We will compare outcome between the NeuroSAFE and standard arm, within the subgroup of participants where bilateral nerve sparing was not recommended.

We will estimate the difference in the primary (IIEF-5 at 12 month) and the main secondary (ICIQ at 3 months) outcomes between arms in this subgroup, using the same analysis model specified in section 4.7 and 4.8 (linear regression adjusted for age, site and baseline IIEF-5 or ICIQ) but restricted to participants without BNSR.

We may also report (without formal testing) on the difference in other secondary and descriptive outcomes within this subgroup, if appropriate.

Although this subgroup analysis may be less powered than the primary analysis of the trial, we expect around 2/3 of the trial participants to belong to this subgroup. This sample size should be sufficient power to detect moderate or large difference between arms.

We will not compare the outcomes within the converse subgroup (participants for whom bilateral never sparing was recommended), which would be of lesser relevance and likely to lack statistical power.

4.11 Missing data sensitivity analysis

We will conduct sensitivity analysis under different missing data assumptions for the primary outcome.

4.11.1 Missing at random

If important predictors of missing data are identified, we may report additional results under missing at random assumption, adjusting for these factors in the main regression model (for baseline predictors), or conducting multiple imputation.

4.11.2 Missing not at random

We will also conduct a sensitivity analysis under not-at-random assumption using reference-based multiple imputation (Carpenter, 2013). We will impute the missing data under the 'last-mean-carried-forward' approach, corresponding to assuming the IIEF-5 score would have remained similar to the last score observed before dropping out (akin to the 'Last-Observation-Carried-Forward' approach, but appropriately allowing for other predictors, and variation in the imputed values).

The multiple imputation model will include the outcomes at each time point, the baseline covariates and other important predictors of missing data and will be stratified by arm. We will assume interim missing (missing assessments with later outcome data available) to be missing-at-random. If some participants have no IIEF-data available (at visit 2, 3 and 4), we will assume they had the lowest score at 3 months (1 for IIEF-5) at visit 2, before conducting the imputation.

The analysis model will be the same as the primary analysis model, and the results will be combined using Rubin's rules.

5 References

Avery, K. a. (2004). ICIQ: a brief and robust measure for evaluating the symptoms and impact of urinary incontinence. *Neurourology and Urodynamics: Official Journal of the International Continence Society*, 322--330.

Cappelleri, J. a. (2005). The Sexual Health Inventory for Men (SHIM): a 5-year review of research and clinical experience. *International journal of impotence research*, 307--319.

Cappelleri, J. C. (1999). Diagnostic evaluation of the erectile function domain of the International Index of Erectile Function. *Urology*, 346--351.

Carpenter, J. R. (2013). Analysis of longitudinal trials with protocol deviation: a framework for relevant, accessible assumptions, and inference via multiple imputation. *Journal of biopharmaceutical statistics*, 1352--1371.

Dinneen, E. a.-H.-P. (2022). Negative mpMRI rules out extra-prostatic extension in prostate cancer before robot-assisted radical prostatectomy. *Diagnostics*, 1057.

Dinneen, E. a.-M.-H. (2022). NeuroSAFE PROOF: study protocol for a single-blinded, IDEAL stage 3, multi-centre, randomised controlled trial of NeuroSAFE robotic-assisted radical prostatectomy versus standard robotic-assisted radical prostatectomy in men with localized prostate cancer. *Trials*, 584.

Gnanapragasam, V. J. (2016). Improving clinical risk stratification at diagnosis in primary prostate cancer: a prognostic modelling study. *PLoS medicine*, e1002063.

Haglind, E. a.-E. (2015). Urinary incontinence and erectile dysfunction after robotic versus open radical prostatectomy: a prospective, controlled, nonrandomised trial. *European urology*, 216--225.

Hays, R. D. (1993). The rand 36-item health survey 1.0. *Health economics*, 217--227.

Hays, R. D. (2001). The RAND-36 measure of health-related quality of life. *Annals of medicine*, 350--357.

Herdman, M. a. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of life research*, 1727--1736.

Leurent, B. a. (2020). Reference-based multiple imputation for missing data sensitivity analyses in trial-based cost-effectiveness analysis. *Health economics*, 171--184.

Rhoden, E. a. (2002). The use of the simplified International Index of Erectile Function (IIEF-5) as a diagnostic tool to study the prevalence of erectile dysfunction. *International journal of impotence research*, 245--250.

Rosen, R. C. (1999). Development and evaluation of an abridged, 5-item version of the International Index of Erectile Function (IIEF-5) as a diagnostic tool for erectile dysfunction. *International journal of impotence research*, 319--326.

StataCorp. (2023). Stata Statistical Software: Release 18. *College Station, TX: StataCorp LLC*.

Uren, A. D. (2020). The International Consultation on Incontinence Questionnaires (ICIQ): An update on status and direction. *Neurourology and Urodynamics*, 1889--1896.

Ware, J. a. (1994). SF-36 physical and mental health summary scales. *A user's manual*.

White, I. R. (2005). Adjusting for partially missing baseline measurements in randomized trials. *Statistics in medicine*, 993--1007.

6 Appendix

6.1 IIEF-5 scoring

The primary outcome for this project is the simplified International Index of Erectile Function (IIEF-5).

The IIEF-5 score, derived from the 15-item International Index of Erectile Function (IIEF), provides a concise evaluation tool for diagnosing erectile dysfunction (ED) and assessing its severity. Rosen selected these five specific items based on their ability to discern the presence or absence of erectile dysfunction, aligning with the NIH definition of the condition (Rosen, 1999). These items primarily centre on erectile function and satisfaction with intercourse.

The questionnaire used was based the British Association of Urological Surgeons (BAUS) version of the IIEF questionnaire, which varies slightly from the original Rosen wording by allowing for a "0- Did not attempt intercourse" response for questions 2, 4, 5 and 7, as described in the table below:

NEUROSAFE PROOF STUDY QUESTIONNAIRE		ROSEN 1999	
Q2	Over the past 3 months when you had erections with sexual stimulation, how often were your erections hard enough for penetration? 0 No sexual activity 1 Almost never or never 2 A few times (less than half the time) 3 Sometimes (about half the time) 4 Most times (more than half the time) 5 Almost always or always	**Q2**	When you had erections with sexual stimulation, how often were your erections hard enough for penetration? 1 Almost never /never 2 A few times (much less than half the time) 3 Sometimes (about half the time) 4 Most times (much more than half the time) 5 Almost always /always
Q4	Over the past 3 months during sexual intercourse, how often were you able to maintain your erection after you had penetrated (entered) your partner? 0 Did not attempt intercourse 1 Almost never or never 2 A few times (less than half the time) 3 Sometimes (about half the time) 4 Most times (more than half the time) 5 Almost always or always	**Q3**	During sexual intercourse, how often were you able to maintain your erection after you had penetrated (entered) your partner? 1 Almost never / never 2 A few times (much less than half the time) 3 Sometimes (about half the time) 4 Most times (much more than half the time) 5 Almost always / always
Q5	Over the past 3 months during sexual intercourse, how difficult was it to	**Q4**	During sexual intercourse, how difficult was it to maintain your erection to completion of intercourse?

	maintain your erection to completion of intercourse? 1 Did not attempt intercourse 2 Very difficult 3 Difficult 4 Slightly difficult 5 Not difficult		 1 Extremely difficult 2 Very difficult 3 Difficult 4 Slightly difficult 5 Not difficult
Q7	Over the past 3 months when you attempted sexual intercourse, how often was it satisfactory for you? 0 Did not attempt intercourse 1 Almost never or never 2 A few times (less than half the time) 3 Sometimes (about half the time) 4 Most times (more than half the time) 5 Almost always or always	**Q5**	When you attempted sexual intercourse, how often was it satisfactory for you? 1 Almost never / never 2 A few times (much less than half the time) 3 Sometimes (about half the time) 4 Most times (much more than half the time) 5 Almost always / always

For the item 5, the "Extremely difficult" response was missing, when it should have been offered as a possible answer and scored as 1. And the "Did not attempt intercourse" answer was scored as 1 when it should have been scored as 0.
To improve comparability, we adjusted the response based on the other IIEF-5 responses. For those who answered "1- "Did not attempt intercourse" and had responded "0 Did not attempt intercourse" to all the questions 2, 4 and 7, we assumed the answer would have been 0 ("Did not attempt intercourse"). For all other participants, we left the score as 1 (corresponding to "Extremely difficult")

The score to the 5 items were then summed, giving an IIEF-5 score ranging from 1 to 25.

6.2 IIEF-6 scoring

IIEF-6 was scored similarly to the IIEF-5. It was based on summing the score to the questions 1, 2, 3, 4, 5 and 15 from the IIEF-15. Q2, Q4 and Q5 are shown in the IIEF-6 table below, and Q1 and Q15 are shown below.

Question 5 was scored as described above, before summing all the item, giving a total score between 1 and 30.

NEUROSAFE QUESTIONNAIRE

Q1	Over the past 3 months how often were you able to get an erection during sexual activity? 0 No sexual activity

	1 Almost never/never 2 A few times (much less than half the time) 3 Sometimes (about half the time) 4 Most times (much more than half the time) 5 Almost always/always
Q15	Over the past 3 months how do you rate your confidence that you could get and keep an erection? 1 Very low 2 Low 3 Moderate 4 High 5 Very high

6.3 ICIQ scoring

	ICIQ
Q3	How often do you leak urine? 0 never 1 about once a week or less often 2 two or three times a week 3 about once a day 4 several times a day 5 all the time
Q4	We would like to know how much urine you think leaks. How much urine do you usually leak (whether you wear protection or not)? 0 none 2 a small amount 4 a moderate amount 6 a large amount
Q5	Overall, how much does leaking urine interfere with your everyday life? Please ring a number between 0 (not at all) and 10 (a great deal)

6.4 Gleason score and Gleason grade group

The Gleason grade is a pathologist's rating of the most predominant pattern and the second most predominant pattern of a biopsy. A patient's Gleason score can be derived by adding these two Gleason grades together. Gleason scores range from 6 to 10, with 6 being the lowest grade of cancer (Gnanapragasam, 2016). The breakdown of the Gleason grade is shown in the table below:

Grade group	Gleason score
GRADE GROUP 1	Gleason score <= 6
GRADE GROUP 2	Gleason score 7 (3+4)
GRADE GROUP 3	Gleason score 7 (4+3)
GRADE GROUP 4	Gleason score 8 (4+4, 3+5, 5+3)
GRADE GROUP 5	Gleason score 9 – 10 (4+5, 5+4, 5+5)

6.5 Cambridge prognostic groups

Cambridge prognostic groups are specified based on the Grade Group or Gleason score, the Prostate specific antigen (PSA) level and the Tumour stage (the T stage from the TNM staging).

Cambridge Prognostic Group 1 (CPG 1)

- a Gleason score of 6, Grade Group 1.
- **and** a PSA level less than 10 nanograms per millilitre (ng/ml)
- **and** a T stage of 1 or 2

Cambridge Prognostic Group 2 (CPG 2)

- [a Gleason score of 3 + 4 = 7, Grade Group 2
- **or** a PSA level between 10 and 20 ng/ml]
  And
- [a T stage of 1 or 2]

Cambridge Prognostic Group 3 (CPG 3)

- [a Gleason score of 3 + 4 = 7, Grade Group 2
- **and** a PSA level between 10 and 20 ng/ml
- **and** a T stage of 1 or 2]
  Or
- [a Gleason score 4 + 3 = 7, Grade Group 3
- **and** a T stage of 1 or 2]

Cambridge Prognostic Group 4 (CPG 4)
If the patient has **one of the following**:

- a Gleason score of 8, Grade Group 4
- PSA level higher than 20 ng/ml
- T stage of 3

Cambridge Prognostic Group 5 (CPG 5)
[If the patient has **two or more of the following**:

- a Gleason score 8, Grade Group 4
- PSA level higher than 20 ng/ml
- T stage of 3]
  Or
- [a Gleason score 9 to 10, Grade Group 5]
  Or
- [T stage of 4]