SURGICAL INTERVENTION TRIALS UNIT

Centre for Healthcare Randomised Trials

REINFORCE

Real-World Evaluation of Robot-Assisted Surgical Services

Statistical Analysis Plan

A UK Collaborative Trial funded by the National Institute of Health Research Health Services and Delivery Research (HS&DR) Programme

Version:	1
Version Date:	09/07/2025
Based on	
Protocol Version:	3.0
Protocol Date:	20/03/2025

NDORMS

NUFFIELD DEPARTMENT OF ORTHOPAEDICS, RHEUMATOLOGY AND MUSCULOSKELETAL SCIENCES

UNIVERSITY OF OXFORD

Investigators

Co-Chief Investigator	
Name:	Professor David Beard
Co-Chief Investigator	
Name:	Professor Marion Campbell

Funder

Name:	National Institute of Health Research Health Services and Delivery Research (HS&DR) Programme
Funder reference number:	NIHR131537
Funder start date:	01/01/2022
Funder end date (anticipated):	31/12/2025

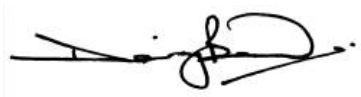
Other

ISRCTN number	18320267

Signatures

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

Co-Chief Investigator

Professor David Beard 

Date:

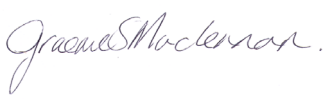
10.07.25

Co- Chief Investigator

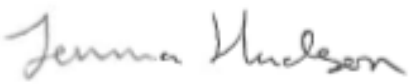
Professor Marion Campbell 

Date: **10.07.25**

Clinical Trials Unit Director

Professor Graeme MacLennan 

Date: **10.07.2025**

Trial Statistician [author] 

Jemma Hudson

Date: **10.07.2025**

Version History

SAP version	Protocol version	Section number changed	Description of and reason for change	Date of change
Version 1	Version 3.0		New document	

Glossary of Abbreviations

ACF	Autocorrelation Function
ARIMA	Autoregressive Integrated Moving Average
CI	Confidence Interval
IPTW	Inverse Probability of Treatment Weighting
ITS	Interrupted Time Series
PACF	Partial Autocorrelation Function
PROM	Patient Reported Outcome Measure
RAS	Robotic Assisted Surgery
SAP	Statistical Analysis Plan
SD	Standard Deviation

Table of Contents

1. Introduction

This statistical analysis plan (SAP) documents the analysis for the REINFORCE study. The SAP is based on the protocol version 3.0 and any deviations from the plan will be described.

Robotic Assisted Surgery (RAS) has the potential to provide high performance and increased access to services for the population, particularly for cancer surgery and other effective high-volume procedures such as arthroplasty; however, the impact of RAS on healthcare service provision is unclear.

2. Study Aims and Objectives

The study aims to undertake a real-world, large-scale evaluation of the introduction and scale-up of RAS services evaluating its impact on NHS service delivery, clinical effectiveness, budget and cost-effectiveness. Specific objectives are to provide evidence on the:

1. Impact of RAS system transformation on clinical and service delivery
2. Potential benefit and harms of RAS across and within speciality areas

Other study objectives that are not in the scope of this SAP:

3. Budget impact and cost-effectiveness to the NHS of the introduction of RAS at scale (not in the scope this SAP). These are described in the protocol and a Health Economics Analysis Plan
4. Mechanisms of change underpinning any change in outcome, including surgeon training (not in the scope this SAP). These are described in the protocol and a Process Evaluation Analysis Plan.

3. General Study Design

REINFORCE was originally planned as a multi-centre pan-specialty stepped-wedge evaluation (with integrated process evaluation and economic assessment, neither are discussed further in this SAP and details can be found in separate companion Health Economic and Process Evaluation Analysis Plans) of the introduction and scale-up of RAS services in the NHS. However, due to the complexities of the roll-out of the intervention in the NHS the design is a multiple interrupted time series (ITS) – this alternative approach was pre-identified as a possibility in the original application and protocol. The design will have a pre and post intervention (change in service delivery) phase.

4. Intervention to be evaluated

The intervention is the introduction or expansion (scaling up to new specialities/new procedures) of RAS services. Sites will either be naïve to RAS services and introducing RAS for the first time (Type 1), will be expanding RAS services to a new specialty (Type 2) or expanding RAS services within a specialty to a new procedure (Type 3).

Each site will identify an index procedure (within a specified specialty) to track the impact of the change in RAS services.

5. Randomisation

Due to the use of ITS design, there is no randomisation of patients or sites.

6. Outcome Measures

Patient-level/clinical outcomes:

- Complications (as measured by the Clavien-Dindo measure of surgical complications)
- Peri- and post-operative complication frequency
- Patient benefit (as measured by the procedure-specific patient reported outcome [specialty specific PROM and normalised across specialties] at 3 months – see below)
- Length of stay
- Time to recovery to normal activities – one question from the EQ-5D-5L
- Quality of life (as measured by EQ-5D-5L)
- Patient experience

Organisational-level outcomes:

- Re-admission rates

Surgeon/team outcomes:

- Operative control
- Operative visualisation

PROMs will be normalised onto a 0-100 scale ahead of the study analysis using the following formula:

$$\text{Transformed scale} = \left[\frac{\text{Actual raw score - lowest possible raw score}}{\text{Possible raw score range}}\right] \text{x100}$$

For the sample size calculation (see below), we used patient benefit as the “primary” outcome, as they generally require the highest sample size to detect meaningful change accepting that outcomes at each level are equally important.

7. Sample Size and Power Calculation

The sample size of the study was based in an incomplete stepped-wedge evaluation with 16 sites over 24 months (the original planned design) and not an interrupted time series design. Each step was to last 4 months, the incomplete design included a 4-month transition to allow sites time to move from pre-RAS-augmented service, and a baseline step where all 16 are control sites. At each step four sites would change. We conservatively assumed that each site would have 8 eligible procedures a month, (32 in each four-month step), giving an available sample size of 2,560 procedures. Assuming an intra-cluster correlation coefficient of 0.01 (appropriate for continuous measures of patient outcome)[1], this sample size had over 90% power to detect 0.3 standard deviations (SDs) and just over 80% power to detect 0.25 SDs difference between arms. Eight procedures a month equates to an average of 2 procedures per week per site. Data on the numbers of patients undergoing likely index procedures suggested that this sample size was fully feasible. For example, the British Association of Urological Surgeons audit data suggests that there are over 7,000 radical prostatectomies undertaken each year in England[2]. A 0.3 SD change represents a “small” to “moderate” change in outcomes[3] – this size of change has been found to be clinically important across a number of settings.

8. Statistical Methods for Objective 1

8.1. General methods

The main analyses will explore the impact of the introduction or scale-up of RAS on service delivery. Final analysis will take place after full recruitment and follow-up. Baseline and follow-up data will be summarised using the appropriate descriptive statistics and graphical summaries and by pre- and post- the introduction or scale up of RAS services. There will be no adjustment to outcomes for multiple testing. All eligible participants will be included in the analysis.

We will use the following nomenclature to describe services before and after the introduction or scale up of RAS services at sites:

- “**pre-RAS**” – this describes the service at each site before the introduction of RAS (site type 1) or scale up of RAS (to a new speciality in site type 2, or to a new procedure in site type 3).

- “**post-RAS**” – this describes the service at each site after the introduction of RAS (site type 1) or scale up of RAS (to a new speciality in site type 2, or to a new procedure in site type 3).

Where possible, secondary analysis will also be done by index procedure type, speciality, by site type and type of robotic system.

8.2. Definition of transition from pre to post RAS services

The transition from pre-RAS to post-RAS will be defined as follows:

- If sites/centres have at least 3 months of data (from site opening) with no robotic procedures reported or included, the transition point from pre-RAS to post-RAS will be defined as from whenever the first robotic procedure occurred,

- If sites/centres have occasional robotic procedures logged in the first 3 months of data collection (from site opening), the change from pre-RAS to post-RAS will be defined at a time point 3 months from the first recorded RAS procedure.

8.3. Interrupted time series analysis

Depending on available data an interrupted time series analysis will be undertaken on all outcomes listed in section 6.

To determine whether the series is linear, the raw data will be plotted and visually checked as well as plotting the residuals versus fitted values. The presence of seasonality will be checked visually by plotting the raw data as well as using autocorrelation function (ACF) and the partial autocorrelation function (PACF). To determine whether autocorrelation is present, the ACF, PACF and the Breusch-Godfrey test[4,5] will be used and the ACF, PACF and the augmented Dickey-Fuller test[4] will be used to confirm the series is stationary.

The exact structure of the model used for each analysis will depend on the structure of the data. For example, if autocorrelation is not present then ordinary least squares segmented regression model will be used, if it is present then an Autoregressive Integrated Moving Average (ARIMA) model will be used. Meta-analysis approaches will be used to pool results from multiple sites.

The proposed specification of the model is:

$$Y_t = \beta_0 + \beta_1 T_t + \beta_2 X_t + \beta_3 (T_t - T_i) X_t + \epsilon_t$$

where Y_t is the outcome at time *t*; T_t is a continuous variable and is time at time *t* i.e. time since the start of the data series i.e. 1,2, …, *n* where *n* is the last time point; X_t is an indicator for the intervention at time *t* and is a dummy variable coded 0 for the pre-RAS phase and 1 for the post-RAS phase; T_i corresponds to the time point, *i*, when the intervention started; ϵ_t is the error term at time *t*. For PROM outcome, baseline score will also be included. Two main intervention effects will be presented change in level and change in slope along with 95% confidence intervals (CIs).

If a site does not have a post-RAS data (e.g. the robot was not introduced), the site will be excluded from the analysis.

8.4. Interim Analysis

There are no planned interim analyses for efficacy or futility.

8.5. Missing Data

8.5.1. Missing Baseline Data

Data missing at baseline for an individual who receives surgery in the pre-RAS or post-RAS will be reported as such. If required, baseline data will be imputed with centre specific mean for continuous data and missing binary/categorical data will include a missing indicator.

8.5.2. Missing Outcome Data

Our primary analysis will be on observed data only. Sensitivities of treatment effect estimate to missing outcome data will be explored if required for PROMs and Complications (as measured by the Clavien-Dindo measure of surgical complications), for example through multiple imputation.

9. Statistical methods for Objective 2

To assess the potential benefit and harms of RAS across and within speciality areas) additional analyses of RAS cases versus non-RAS cases will also be conducted. These will include all RAS cases versus all non-RAS cases (immaterial of time point); non-RAS cases in the pre-phase vs RAS cases in the post phase; and a restricted analysis of all RAS cases vs all non-RAS cases in the post-phase only. Data will be summarised using the appropriate descriptive statistics. To reduce potential confounding from non-randomised treatment allocation, we use propensity score methods in the regression analysis to balance observed baseline characteristics between RAS and non-RAS groups. Propensity scores will be estimated using logistic regression models including key covariates such as age, sex, ASA grade, speciality and baseline PROMs. Depending on the distribution and overlap, we will apply propensity score matching, inverse probability of treatment weighting (IPTW), or covariate adjustment using the propensity score. These methods will be applied to the main comparisons (e.g., all RAS vs all non-RAS; pre-phase non-RAS vs post-phase RAS; post-phase only), and the resulting adjusted analyses will use logistic or linear regression as appropriate. Sensitivity analyses will be conducted to assess the robustness of findings across propensity score techniques. Treatment effects will be summarised with 95% CIs.

10. Statistical software

All analysis will be carried out in Stata 18[6] (or the current version available).

11. Dummy Tables

Table 1. Number of participants recruited

Site	Pre-RAS N=	Post-RAS N=	Total N=
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			

Table 2. Number of patients recruited in each period

Site	Period - N							
	1	2	3	4	5	6	7	8
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								

Period shading: yellow shading represents pre-RAS; green shading represents post-RAS. Each period will represent a time period (e.g. one month).

Table 3. Participant characteristics[1]

Characteristics	Pre-RAS N=	Post-RAS N=
Sex – n (%)		
Male		
Female		
Prefer not to say		
Age (years) – mean (SD); n		
Ethnicity – n (%)		
White		
Mixed/multiple ethnic groups		
Asian/Asian British		
Black/African/Caribbean/Black British		
Other		
Specialty – n (%)		
Urology		
Colorectal		
Gynaecology		
Orthopaedics		
Upper Gastro-intestinal		
HPB		
Renal		
EQ-5D – 5L – mean (SD); n		
EQ-5D VAS – mean (SD); n		
Specialty specific PROMS – mean (SD); n		
PROMS 1		
PROMS 2		
PROMS 3		
PROMS 4		
PROMS 5		
PROMS 6 etc.		

SD Standard Deviation

[1]This table will also be presented by Non-RAS participants vs RAS participants

Table 4. Operative details – overall, index procedure type, specialty, site type and type of robotic system [1]

Details	Pre-RAS N=	Post-RAS N=
Number of participants that had surgery – n (%)		
Procedure – n (%)		
ASA grade assessment by anaesthetist – n (%)		
Grade I		
Grade II		
Grade III		
Grade IV		
Planned operation type – n (%)		
Robot-assisted		
Minimally invasive		
Open		
Unplanned conversion – n (%)		
Yes		
No		
Operation type for those who has an unplanned conversion – n (%)		
Open		
Laparoscopic		
Robot-assisted		
Other		
Reason for unplanned conversion – n (%)		
No trained staff available		
Robot not working		
Other		
Surgery time – mean (SD), median [IQR]; n		
Time from anaesthetic room to time out of recovery – median [IQR]; n		
Recovery time – mean (SD), median [IQR]; n		
Level of surgeon performing majority of the case – n (%)		
Consultant		
Fellow		
Registrar		
Equipment failure – n (%)		
Yes		
No		
Reason for equipment failure – n (%)		
Equipment malfunction (robotic)		
Equipment malfunction (non-robotic)		
Other		
Admitted to ICU[2] – n (%)		
Yes		
No		
How many times admitted to ICU – n (%)		
One		
Two		

Etc.	
ICU length of stay (total) – mean (SD), median [IQR]; n	
ICU length of stay (first admission)	
ICU length of stay (second admission)	
ICU length of stay (etc admission)	
Admitted to HDU[3]	
Yes	
No	
How many times admitted to HDU – n (%)	
One	
Two	
Etc.	
HDU length of stay – mean (SD), median [IQR]; n	
HDU length of stay (first admission)	
HDU length of stay (second admission)	
HDU length of stay (etc admission)	

IQR Interquartile range
[1]This table will also be presented by Non-RAS participants vs RAS participants
[2] of the # admitted to ICU, # were also complications
[3] of the # admitted to HDU, # were also complications

Table 5. Reasons for not having surgery– overall, index procedure type, specialty, site type and type of robotic system [1]

	Pre-RAS N=	**Post-RAS N=**
Number of participants that did not have surgery – n (%)		
Reasons for not having surgery – n (%)		
Patient unwell		
Patient declined surgery		
Other		
Number of cancellations – n (%)		
One		
Two etc		
Who cancelled the operation (by number of cancellations)– n (%)		
Patient		
Hospital		
Clinician		

[1]This table will also be presented by Non-RAS participants vs RAS participants

Table 6. Clinical outcomes - overall, index procedure type, specialty, site type and type of robotic system [1]

	Pre-RAS N=	Post-RAS N=
Number of participants with a peri- and post-operative complication – n (%)		
Number of peri- and post-operative complications - n		
One		
Two		
Three		
etc		
Peri-operative complications		
Number of participants with a complication – n (%)		
Number of complications - n		
One		
Two		
Three		
Details – n		
Unplanned transfusion		
Major haemorrhage protocol activated		
Injury to organ		
Anaesthetic event		
Other		
Post-operative complications		
Number of participants with a complication – n (%)		
Number of complications - n		
One		
Two		
Three		
Clavien-Dindo profile		
Grade - n		
Grade I		
Grade II		
Grade III		
IIIa		
IIIb		
Grade IV		
IVa		
IVb		
Grade V (mortality)		
Details - n		
Cardiac		
Respiratory		
Neurological		
Gastrointestinal		
Renal		
Other		
Length of stay – mean (SD); n		
Time to recovery to normal activities – mean (SD); n		

EE Effect estimate; SD standard deviation; CI confidence interval
[1]This table will also be presented by Non-RAS participants vs RAS participants

Table 7. Participant level outcomes - overall, index procedure type, specialty, site type and type of robotic system [1]

	Pre-RAS N=	Post-RAS N=
EQ-5D-5L – mean (SD); n		
Baseline		
3 months		
EQ-5D-5L VAS – mean (SD); n		
Baseline		
3 months		
Specialty specific PROMS – mean (SD); n		
PROMS 1		
Baseline		
3 months		
PROMS 2		
Baseline		
3 months		
PROMS 3		
Baseline		
3 months		
PROMS 4		
Baseline		
3 months		
PROMS 5		
Baseline		
3 months		
PROMS 6		
Baseline		
3 months		
Patient experience		
Involvement in decisions		
Yes, definitely		
Yes, to some extent		
No		
Information about condition or treatment provided		
Too much		
The right amount		
Not enough		
No information was given		
Explanation about operation or procedure including risks and benefits		
Yes, completely		
Yes, to some extent		
No		
I did not want an explanation		

Questions about operation or procedure answered in an understanding way	
Yes, completely	
Yes, to some extent	
No	
I did not have any questions	
How the operation or procedure was explained in an understanding way	
Yes, completely	
Yes, to some extent	
No	
Don't know/ can't remember	
Would you recommend operation	
Yes, definitely	
Yes, to some extent	
No	
Don't know	
Rate of care overall	
Excellent	
Very good	
Good	
Fair	
Poor	

EE Effect estimate; SD standard deviation; CI confidence interval
[1]This table will also be presented by Non-RAS participants vs RAS participants

Table 8. Organisation-level outcomes re-admission rates - overall, index procedure type, specialty, site type and type of robotic system [1]

	Pre-RAS N=	Post-RAS N=
Number of participants re-admitted to hospital – n (%)		
Number of re-admissions – n (%)		
One		
Two etc.		

EE Effect estimate; CI confidence interval.
[1]This table will also be presented by Non-RAS participants vs RAS participants

Table 9. Surgeon questionnaire - overall, index procedure type, specialty, site type and type of robotic system [1]

	Pre-RAS N=	Post-RAS N=
Able to visualise the area of interest – n (%)		
Not at all		
A little		
Quite a bit		
Very well		
Number of procedure types done for the surgeon who did majority of the cases – n (%)		
< 20		
21-50		
51-100		
> 100		
Number of techniques done for the surgeon who did majority of the cases – n (%)		
< 20		
21-50		
51-100		
> 100		
SURG-TLX - mean (SD); n		
Mental demands		
Physical demands		
Temporal demands		
Task complexity		
Situational stress		
Distractions		

EE Effect estimate; SD standard deviation; CI confidence interval

[1]This table will also be presented by Non-RAS participants vs RAS participants

Table 10 Interrupted time series analysis - overall, index procedure type, specialty, site type and type of robotic system [1]

	Change in level			Change in slope		
	Estimate	**95% CI**	**p-value**	**Estimate**	**95% CI**	**p-value**
Complications measured by Clavien-Dindo						
Peri- and post-operative complication						
Specialty specific PROMS						
PROMS 1						
PROMS 2						
PROMS 3						
PROMS 4						
PROMS 5						
PROMS 6						
Length of stay						
Time to recovery to normal activities						
EQ-5D-5L						
EQ-5D-5L VAS						
Patient experience						
Involvement in decisions						
Information about condition or treatment provided						
Explanation about operation or procedure including risks and benefits						
Questions about operation or procedure answered in an understanding way						
How the operation or procedure was explained in an understanding way						
Would you recommend operation						
Rate of care overall						
Re-admission rates						
Operative control						
Mental demands						
Physical demands						
Temporal demands						

	Task complexity
	Situational stress
	Distractions
Operative visualisation	

CI confidence interval

Table 11 RAS versus no RAS analysis - overall, index procedure type, specialty, site type and type of robotic system [1]

	Estimate	95% CI	p-value
Complications measured by Clavien-Dindo			
Peri- and post-operative complication			
Specialty specific PROMS			
PROMS 1			
PROMS 2			
PROMS 3			
PROMS 4			
PROMS 5			
PROMS 6			
Length of stay			
Time to recovery to normal			
EQ-5D-5L			
EQ-5D-5L CAS			
Patient experience			
Involvement in decisions			
Information about condition or treatment provided			
Explanation about operation or procedure including risks and benefits			
Questions about operation or procedure answered in an understanding way			
How the operation or procedure was explained in an understanding way			
Would you recommend operation			
Rate of care overall			
Re-admission rates			
Operative control			
Mental demands			
Physical demands			
Temporal demands			
Task complexity			
Situational stress			
Distractions			
Operative visualisation			

CI confidence interval

12. Dummy Figures

Interrupted time series analysis

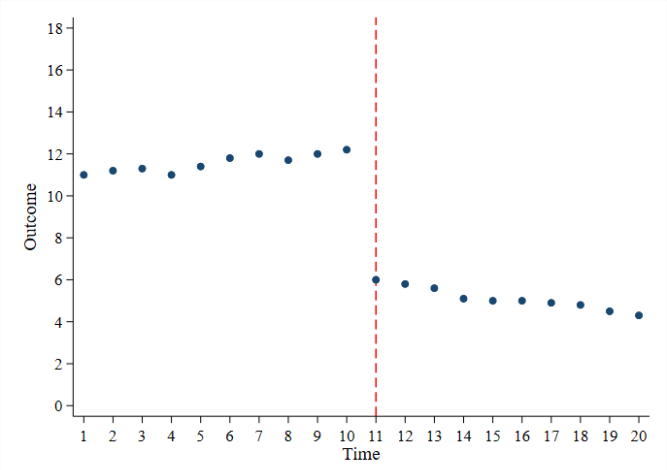


Figure 1. Example figure for showing outcome pre- and post- RAS – overall, index procedure type, specialty, site type and type of robotic system [1]
Red line represents the implementation of pre-RAS to post-RAS

13. References

1. Cook JA, Bruckner T, MacLennan GS, Seiler CM. Clustering in surgical trials - database of intracluster correlations. *Trials*. 2012;13(1):2. https://doi.org/10.1186/1745-6215-13-2. doi: 10.1186/1745-6215-13-2.

2. BAUS. British association of urological surgeons: BAUS data & audit programme annual Report2019. .

3. Cohen J. *Statistical power analysis for the behavioral sciences.* New York: Academic Press; 1977.

4. Godfrey LG. Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica*. 1978;46(6):1293–1301. http://www.jstor.org/stable/1913829. doi: 10.2307/1913829.

5. BREUSCH TS. Testing for autocorrelation in dynamic linear models*. *Aust Econ Pap*. 1978;17(31):334–355. https://doi.org/10.1111/j.1467-8454.1978.tb00635.x. doi: 10.1111/j.1467-8454.1978.tb00635.x.

6. StataCorp. 2023. stata statistical software: Release 18. college station, TX: StataCorp LLC.