Surgery or Cast for Injuries of the EpicoNdyle in Children's Elbows (SCIENCE) Study

Statistical Analysis Plan

Version 1.0– 31 Oct 2024

Linked to File Note[1]

Based on Protocol version 6.0 – 19 Aug 2021

Trial registration: ISRCTN16619778

Oxford Clinical Trials Research Unit (OCTRU)
Centre for Statistics in Medicine (CSM)

[1] File Note on SAP Data Definition and Tables can be found in: Z:\CSM_SCIENCE\Statistical Analysis\5. Statistical Analysis Plan (SAP)\Current\ SCIENCE_FileNote_SAPDD&T_28Oct2024.docx

CONTENTS

1. INTRODUCTION

This document details the proposed data presentation and analysis for the main paper(s) and final study reports from the **National Institute of Health Research Health Technology Assessment (HTA) funded the multicentre randomised controlled superiority trial of operative fixation versus non-operative treatment for medial epicondyle fractures of the humerus in children (SCIENCE).** The results reported in these papers should follow the strategy set out here. Subsequent analyses of a more exploratory nature or of extended study follow-up will not be bound by this strategy unless explicitly stated to be covered, though they are expected to follow the broad principles laid down here. The principles are not intended to curtail exploratory analysis (for example, to decide cut-points for categorisation of continuous variables), nor to prohibit accepted practices (for example, data transformation prior to analysis), but they are intended to establish the rules that will be followed, as closely as possible, when analysing and reporting the trial. This document follows published guidelines regarding the content of statistical analysis plans for clinical trial [1, 2].

The analysis strategy will be available on request when the principal papers are submitted for publication in a journal. Suggestions for subsequent analyses by journal editors or referees, will be considered carefully, and carried out as far as possible in line with the principles of this analysis strategy. If reported, the analyses will be marked as post-hoc; the source of the suggestion will be acknowledged, and the reader will be advised to rely primarily on the pre-specified analysis for the interpretation of the results.

Any deviations from the statistical analysis plan will be described and justified in the final report of the trial. The analysis should be carried out by an identified, appropriately qualified and experienced statistician, who should ensure the integrity of the data during their processing. Examples of such procedures include quality control and evaluation procedures.

At the time the trial setup no SAP Data Definitions and Tables (SAP DD&T) were mandated by the OCTRU SOPs. The related SAPDD&T documentation is not available in this trial in that form. Examples of Shell tables are documented in the previous ISAP document found here:

Z:\CSM_SCIENCE\Statistical Analysis\3. ISAP\ISAP Superseded Final Versions\SCIENCE_ISAP_V1.0_03June2019.docx. The ISAP will provide the basis for the Shell Tables for the final analyses report and will be reviewed prior to the start of the final analysis and any important changes and reasons will be documented in the report.

1.1 Key personnel

Author(s): Elnaz Saeedi (Trial Statistician), Nicholas Peckham (Senior Statistician), Daphne Kounali (Lead Statistician)

Reviewers: Ioana Marian (Senior Statistician)

Approver: Daphne Kounali (Lead Statistician), Daniel Perry (Chief Investigator)

1.2 Changes from previous version of SAP

This is the first version of the SAP. Therefore, there are no previous versions listed in the table below which gives a summary of key changes from any earlier versions of SAP. If this SAP is updated a summary of changes will be added with particular relevance to protocol changes that have an impact on the design, definition, sample size, data quality/collection or analysis of the outcomes below.

Version number Issue date	Author of this issue	Protocol Version & Issue date	Significant changes from previous version together with reasons
Not applicable		Not applicable	Not applicable as this is the 1st issue

2. BACKGROUND AND OBJECTIVES

The management of fractures of the medial epicondyle is a controversial issue in paediatric fracture care [3]. These fractures typically occur in children around 10-12 years old [4], with or without dislocation of the elbow joint. The debate for clinicians is whether to realign and hold the bone fragments with operative fixation, or whether to allow the fragments to heal in their current position without surgery by resting the elbow in a cast.

Observational studies have demonstrated support for both operative and non-operative treatment strategies, generating uncertainty amongst surgeons. Two published systematic reviews have demonstrated disagreement in the management of this injury [4, 5]. One systematic review concluded that nonsurgical treatment offers excellent functional results equivalent to surgical treatment [5]. The other concludes that surgical fixation should be strongly considered to achieve union of the bone fragments thereby maximising elbow stability for an increasingly athletic child population [4]. To add further to the debate, a widely used 'evidence-based review' textbook has recently advocated against surgery, citing increased long-term pain and stiffness compared to non-operative treatment [6].

Much of the controversy has arisen because there have been no prospective studies evaluating the treatment of these fractures. Current literature has serious methodological limitations, especially with regard to inconsistent follow-up, no standardisation to treatment approaches, infrequent use of patient reported outcomes, and selection bias for operative fixation [6]. There has also been a lack of agreement on how to record successful outcomes; with radiographic union of the fracture fragments being the most commonly used outcome, and pain or function being infrequently recorded, although there is known to be poor correlation between radiographic union and functional outcomes [5]. The uncertainty within the literature has resulted in considerable variation in clinical practice. There is an increasing tendency toward surgery, which has been particularly driven by US literature identifying the athletic demands of children and adolescents, and the expectations of early mobilisation and return to sport [3-7].

An audit of surgical practice amongst 30 centres in the UK was conducted and identified 520 medial epicondyle fractures over a 3-year period. Overall, 225 (43%) of these fractures were treated with surgical fixation, with the remaining 295 (57%) treated non-operatively. 39 children (8%) had an incarcerated fragment, which is an absolute requirement for surgery. The decision to offer 'operative intervention' is highly dependent upon the surgeon. Surgical fixation is thought to improve the likelihood of 'bony union' of the fracture. However, this is balanced against the small but definite risks from the surgery including infection, nerve damage around the elbow, broken and retained metalwork, and the risks associated with general anaesthesia. It is unclear whether bony union has any bearing on functional recovery, including return to sports.

There is a clear and pressing need to inform patients about the benefits or otherwise of operative fixation versus non-operative treatment, and the need to inform commissioners regarding the costs of the different treatment strategies to the NHS and society. The SCIENCE study is a randomised superiority trial of operative fixation versus non-operative treatment for medial epicondyle fractures of the humerus in children [8, 9].

The aim of this pragmatic randomised controlled superiority trial is to evaluate the clinical and cost-effectiveness of operative fixation versus non-operative treatment for displaced medial epicondyle fractures of the elbow in children.

The estimand for the primary objective (including the analysis of the primary outcome) is described in Table 1.

Table 1: Estimand-to-analysis table

Primary Objective: To quantify and draw inferences on observed differences in function for children between operative fixation versus non-operative fixation at 1-year post-randomisation for fractures of the medial epicondyle in children	
Estimand: A single primary estimand will be the difference in the levels of physical function in children aged 7 to15 years old between those treated with operative and non-operative fixation, 1 year following an injury no more two weeks old, leading to an uncomplicated displaced medial epicondyle fracture of the humerus, irrespective of fixation technique used and any unforeseen technical difficulties making the delivery of either randomised intervention impossible or difficult and/or receiving any other interventions that are part of the standard of care.	
Treatment: Operative fixation vs non-operative treatment	
Estimand	***Analysis***
Target population Children aged between 7 and 15 years who present at participating UK hospitals with a medial epicondyle fracture of the humerus.	***Analysis set*** All randomised participants.
Variable Physical Function measured on the child reported PROMIS Upper Extremity scale at 1-year post-randomisation.	***Outcome measure*** Continuous outcome bounded scale (PROMIS Upper Extremity T-score) where higher scores indicate better function better function, at 12 months post-randomisation
Handling of intercurrent events **Switching Treatment arm:** Participants who receive a different intervention to that which they were allocated will be analysed according to their randomised allocation **Further surgery:** Participants that require additional surgical intervention on their elbows are also considered part of the treatment All intercurrent events will be handled under a treatment-policy strategy [1].	***Handling of missing data*** Multilevel modelling of outcome scores over time, imputes implicitly intermittent missing data over the post-randomisation period of follow-up. The underlying assumption is missing at random (MAR) conditional on all other outcome scores and minimisation factors. This is the main approach to handling missing data. *Sensitivity analysis:* The MAR assumption can be strengthened by conditioning on additional baseline covariates predictive of the outcome and missingness. Missing data arising from participants discontinuing in the study prior to 12 months under a NMAR (Not Missing at random) informative missing data mechanism will be explored under two approaches: a) the extended Hausman-Wise-Diggle-Kenward selection model when drop out depends on outcome scores prior to drop out and where this is jointly modelled with the longitudinal outcome. b) using controlled multiple imputation: δ- based imputation (see section (Details in Section 6.3: Missing data) where the primary analysis model is augmented with an offset term with values set to the fixed values of the missing and repeated over

	a range (+/- 1 SD PROMIS score). This is done to assess the impact of drop-out on inferences
Population-level summary measure	***Analysis approach***
Average Treatment effect (ATE): Mean difference in physical function scores at 12 months, between the two randomised arms, as randomised, measured on the PROMIS Upper extremity scale.	Analysis of repeated measures of PROMIS Upper Extremity scores over time including all randomised participants as randomised and including all measurement occasions (at baseline, and all follow-up times up to 12 months) using multilevel linear random effects model. Sensitivity analyses: Primary estimand (ITT) for the average treatment effect (ATE) using regression adjustment for the propensity of allocated treatment received and including the quantiles of the propensity score distribution as a covariate.

[1]Strategies defined in E9 (R1) include treatment policy, while on treatment, principal stratum and hypothetical [1, 10]

3. STUDY METHODS

3.1 Trial Design/framework

SCIENCE is a multi-centre, prospective, randomised, superiority trial using a two-arm parallel group design in at least 35 centres across the UK, New Zealand and Australia. Children aged 7-15 years presenting for treatment for displaced medial epicondyle fracture of the elbow, will be randomised in a 1:1 ratio to either receive operative fixation, or non-operative treatment. The nature of the treatment means that neither the patients and their parents/guardians, nor the treating clinician can be blinded to the treatment received; however, the treating clinical team will take no part in the follow-up assessments and outcome assessors will be blinded to the participant's treatment allocation.

The trial includes an internal pilot (phase 1) which was designed to confirm the expected rate of recruitment in a large-scale, multi-centre randomised controlled trial. This pilot was planned to take place at 20 centres over 12 months. The Data Safety and Monitoring Committee (DSMC) was tasked to make a recommendation to the Trial Steering Committee (TSC) regarding trial continuation in the event that the recruitment target for the internal pilot is not met. Otherwise, the trial would continue into the main phase (phase 2), and patients from the internal pilot would be included in the final analysis. The main trial phase was planned to be recruiting from a minimum of 35 centres treating children's fractures across the UK, stratified by centre to account for centre specific effects, and by dislocation status of the elbow. Follow-up was planned to be made electronically (sent by e-mail or text message) for the Patient Reported Outcomes at 6 weeks, 3 months, 6 months and 12 months (Primary outcome time-point).

All children aged 7-15 years old presenting at the trial centres with a displaced medial epicondyle fracture of the humerus are potentially eligible to take part in the trial. After consent has been gained, a local research associate will collect baseline demographic data, the Patient Reported Outcomes Measurement Information System (PROMIS) Upper Extremity Score for Children Computer Adaptive Test, DASH S/PA Module, Wong Baker Faces Pain Score, and health-related quality of life using the EuroQoL EQ-5DY.

3.2 Randomisation and Blinding

Patients will be randomised after consent. All hospital treatment areas have access to the internet so will access the randomisation service in real time i.e. there will be no delay in patient treatment.

Consenting participants will be allocated randomly (1:1) to either operative fixation or non-surgical treatment. Randomisation will be performed using a minimisation algorithm to ensure balanced allocation of participants across the two treatment groups, stratified by centre and dislocation status of the elbow at presentation in the emergency department (i.e. dislocated or not dislocated). The first 30 participants will be randomised using a simple randomisation schedule produced by the trial statistician, to seed the minimisation algorithm, and a non-deterministic probabilistic element will be introduced to prevent predictability of the treatment allocation.

Stratification by centre within the minimisation algorithm will help to ensure balanced treatment assignment within centre so that any unobserved factors (e.g. case-mix) contributing to between-centre variation in practices are balanced between treatment arms. The catchment area (the local population served by the hospital) will be similar for all hospitals; each hospital being a children's injury unit dealing with these fractures on a regular basis. All of the recruiting hospitals, and indeed all hospitals throughout the NHS, use these techniques as part of their normal practice i.e. staff will already be equally familiar with both forms of treatment. This cannot eliminate the clinician-specific effect of an individual at any one centre. However, since the procedures are commonplace across the NHS, many clinicians will be involved in the management of this group of patients; likely between 5 and 20 clinicians at each centre, including consultants and trainee surgeons. Therefore, we anticipate that each individual clinician will only treat a handful of those enrolled in the trial, reducing the risk of a clinician-specific effect upon the outcome in any one centre.

Stratification by dislocation-status of the elbow within the minimisation algorithm (i.e. not dislocated at presentation to emergency department, or dislocated at presentation to emergency department with a subsequent satisfactory reduction) within the minimisation algorithm will help to ensure that the perceived severity of the injuries through additional soft-tissue damage are balanced across the treatment groups. Any participants for whom the elbow dislocation cannot be reduced prior to randomisation, will be excluded from the trial.

Full details of the randomisation are available in SCIENCE_RBP_V1.0_08May2019, stored in the confidential statistical section of the TMF. No emergency randomisation plan was developed for this study and no emergency randomisation was carried out.

Patients and their parents/guardians cannot be blinded to their treatment. The treating clinician will of course, not be blinded to the treatment they are providing. However, the treating clinical team will take no part in the follow-up assessment of the patients. The outcome data will be collected directly from the patient and/or their parents/guardians. Outcome assessors will be blinded to the participant's treatment allocation.

3.3 Sample Size

Extremity Score for Children. Raw scores are translated into standardised T scores with a population mean of 50 and a Standard Deviation (SD) of 10. The 'Minimally Clinically Important Difference' (MCID) for the PROMIS Upper Extremity Score amongst children with milder forms of disability has been demonstrated to be three to

four [11], we seek to find a difference of four points. Assuming a SD of 10, a total of 266 patients (133 in each arm) is required to provide 90% power at the 5% level (two-sided) to detect a 4-point difference between the groups. Allowing for a conservative 20% loss to follow-up, we plan to recruit 334 patients in total for this trial (167 per arm).

The SD of 10 derived by PROMIS was based on a sample of children with a higher proportion of chronic illness than the general population [12], and it is anticipated that the variation in outcomes in the treatment of acute medial epicondyle fractures is likely to be less than in a chronic illness.
Independent verification of the sample size is available in Z:\CSM_SCIENCE\Statistical Analysis\4. Sample Size**SCIENCE_SampleSize_Verification_04Apr2019.docx**, stored in the confidential statistical section of the TMF.

3.4 Statistical Interim Analysis, Data Review and Stopping guidelines

The Data and Safety Monitoring Committee (DSMC) is a group of independent experts external to the trial who assess the progress, conduct, participant safety, and, if required, critical endpoints of a clinical trial. There are no formal interim analyses and therefore no formal stopping rules for this study.
The DSMC will follow the charter, based on the DAMOCLES guidelines[13], as described in the document SCIENCE_DSMC_Charter_V2.0_09May2019 stored in the TMF. The DSMC will undertake interim review of the trial's progress including updated figures on recruitment, data quality, adherence to protocol and follow-up, and main outcomes and safety data and review the sample size assumptions as described previously.

They will also consider emerging evidence from other related trials or research and review related SAEs that have been reported. They may advise the chair of the Trial Steering Committee at any time, if, in their view, the trial should be stopped for ethical reasons, including concerns about participant safety. DSMC meetings will be held at least annually during the recruitment phase of the study.

3.5 Timing of Analysis

The final analysis of all primary and secondary endpoints will be conducted together when all recruited patients have reached their planned follow-up up to 12 months and the relevant data received, cleaned and finalised.
The trial also includes long-term follow-up from randomisation until skeletal maturity has been reached by the participants; the analysis of this data will be reported separately following a separate statistical analysis plan.

3.6 Blinded analysis

A blinded analysis of the data (not separated by treatment arm) will be undertaken prior to the final data lock to undertake data cleaning, to look into the distribution of variables, missing data distributions, and to review exclusions associated with the per protocol population.

3.7 Statistical Analysis Outline

It is anticipated that all the statistical analysis will be undertaken using Stata [14] although other well-validated statistical packages may be considered for statistical quality control validation analyses. All analyses will be carried out under a treatment policy strategy (that is, all patients will be analysed in the group they were randomised to regardless of actual treatment received. The analyses will be repeated for the per protocol

population(definition in section 4.2) as mandated by the protocol , bearing in mind that this may introduce bias by losing the benefits of randomisation and introducing ambiguity in the primary estimand of interest.

Although we have allowed for up to 20% missing data in the sample size, we hope to minimise this by utilising data collection techniques appropriate to the age of participating children. Before carrying out the within trial analysis, we will check the trial data for any missing data. Where possible the reasons for missing data will be ascertained and reported. The nature and pattern of the 'missingness' will be carefully considered — including whether data can be treated as missing at random (MAR). The handling missing data is detailed 6.3.

Standard descriptive statistics will be used to describe the demographics between the treatment groups reporting means and standard deviations or medians and interquartile ranges as appropriate for continuous variables, and numbers and percentages for binary and categorical variables. All comparative outcomes will be presented as summary statistics and reported together with 95% confidence intervals and all tests will be carried out at a 5% two-sided significance level.

The Patient Reported Outcomes Measurement Information System (PROMIS) Upper Extremity Score for Children at 12 months is the primary outcome of the study and the primary analysis will compare this between the treatment groups in a linear mixed effects method including all participant data, adjusting for the stratification factors. A simple analysis of covariance (ANCOVA) of the primary outcome at 12 months adjusting only for the baseline PROMIS score will be undertaken as a secondary analysis.

Subgroups analyses will be undertaken with the aim of exploring the consistency of treatment effects across important baseline characteristics. These include age-groups (8+ or younger), initial immobilisation duration and dislocation status will be explored using treatment by subgroup interactions [15]. Forest plots with 95% confidence interval will be used to present the subgroup analysis results. Secondary clinical outcomes and patient reported outcomes will be similarly analysed using mixed effects regression, using logistic regression for binary data and linear regression for continuous data.

Any subsequent changes in this plan are strongly discouraged. If any changes to this SAP are implemented these will be documented in an updated version of this SAP. Documentation of any such changes will include the rationale for the changes, and when and by whom these were agreed.

4. STATISTICAL PRINCIPLES

4.1 Statistical Significance and Multiple Testing

There is a single pre-specified primary outcome, so there is no correction for multiple testing. A significance level of 0.05 will be used, with 95% confidence intervals reported. All secondary analyses will be considered as supporting the primary analysis and will also be analysed using a significance level of 0.05 with 95% confidence intervals.

Interim analyses of primary and secondary outcomes will not be carried out unless requested by the DSMC. In this case, p-values of 0.001 will be used for significance and 99% confidence intervals presented.

4.2 Definition of Analysis Populations

Populations for analysis are defined as follows:

- Intention-to-treat (ITT): all participants analysed in their randomised groups, regardless of actual treatment received. This reflects the primary estimand of interest (treatment policy).
- Per-Protocol (PP): participants who received the intervention as intended will be analysed according to the treatment they actually received. Participants will be excluded from the per-protocol population if:

- They did not receive the treatment allocated through randomisation
- They did not fully satisfy the eligibility criteria for the study

Blinded review of the protocol deviation data (not separated by treatment arm) which may affect the PP analyses will be undertaken prior to the final data lock. A summary of the characteristics of exclusions to define the per-protocol population will be reported. The motivation for any such exclusion is to examine the robustness of inferences to factors that can threaten the study's internal validity. Deviations associated with intercurrent events which occurred in the trial setting but would not be observed in practice are considered the main threat, consistent with the primary ITT estimand. If a significant number of such deviations are present e.g. involving more than >5% of participants, the results of the primary analyses may introduce ambiguity in the interpretation. In this case sensitivity analyses is undertaken and detailed in section 6.4.

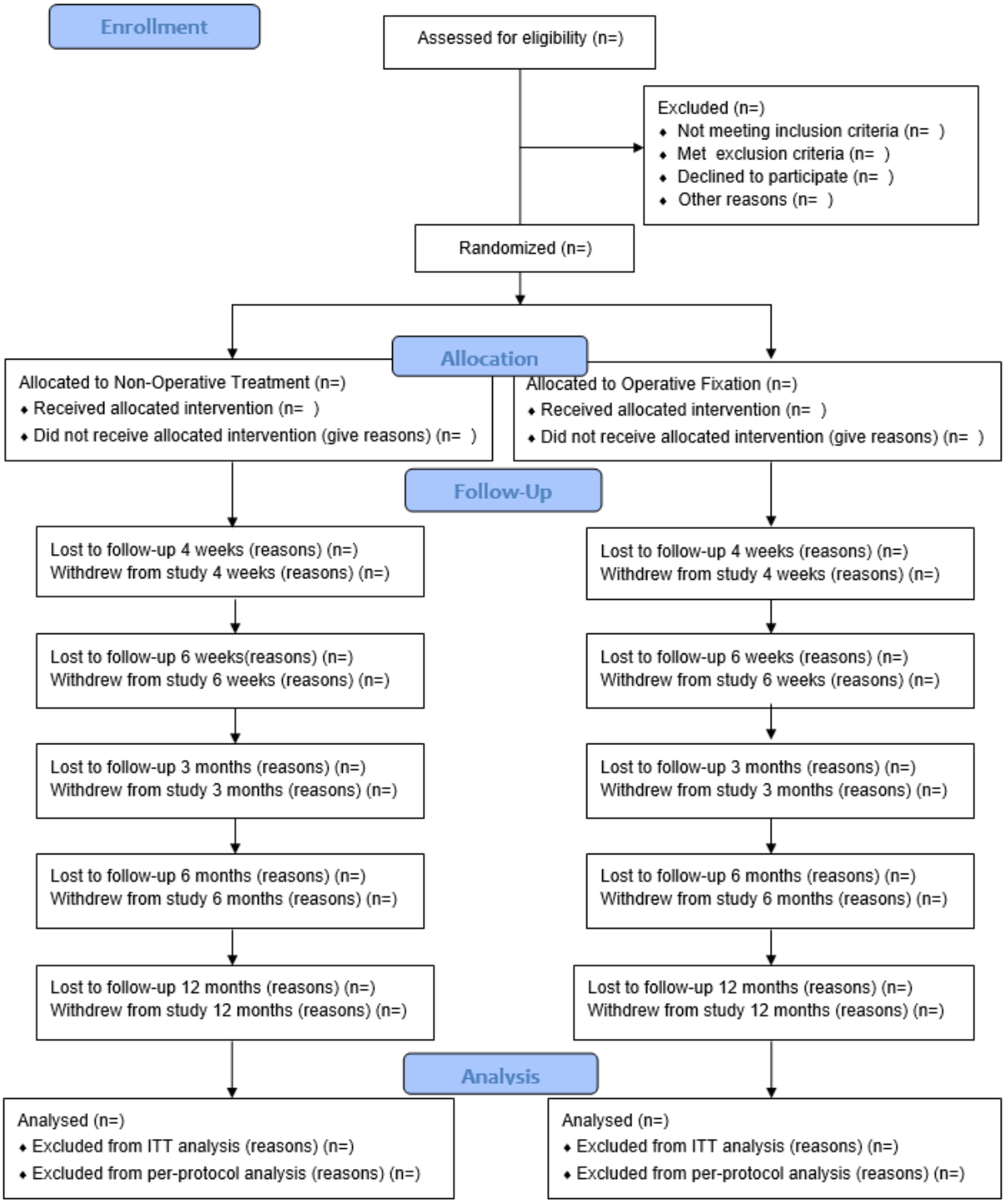
5. TRIAL POPULATION AND DESCRIPTIVE ANALYSES

Summary of flow of trial participants through the trial and baseline stratification, demographic and clinical characteristics of each group.

5.1 Representativeness of Study Sample and Patient Throughput

A CONSORT Flow Chart (Figure 1) will be used to summarise the flow of participants through each stage of the trial, including a breakdown of the number of participants in each stage of the trial from screening through recruitment to the end of the trial. The number participants who are excluded, declined consent, withdrew or were lost to follow-up are also summarised in the flow chart.

Figure 1 The SCIENCE CONSORT flowchart



Additional expansion of reasons and timing of withdrawals from the study will be included as needed as in Table 2 below.

Table 2: Reasons for exclusion

REASON FOR EXCLUSION	Overall DDMonYYYY-DDMonYYYY
Not meeting inclusion criteria	
Injury > 2 weeks old	
Incarceration of medial epicondyle fragment within elbow joint	
Complex elbow fracture (i.e. fracture extending into joint)	
Sustained additional fracture	
Dislocated elbow unable to be realigned into a satisfactory position	
Insufficient English language	
Developmental delay	
Developmental abnormality	
No access to mobile phone/email (internet access)	
Other	
Total Ineligible	
Eligible but refused	
Child does not want to be part of the research project	
Parent/guardian does not want to be part of the research project	
Child does not want to complete the questionnaires	
Parent/guardian does not want to complete the questionnaires	
Treatment Preference	
Non-operative treatment	
Surgical fixation	
Patient decision – no reason given	
Other	
Other	
No legal parental representative present	
Internet problems	
Research staff not available/informed	
Other	
Total eligible but not randomised	
Total patients screened but not randomised	

Protocol deviations will be summarised in Table 3 by treatment, on the impact and the importance of deviations.

Table 3 Details of protocol deviations

	Non-Operative Treatment (N =)	Operative fixation (N =)	Total (N =)
Number of protocol deviations			
Primary outcome available			
Protocol deviation impact			
Completeness of trial data			
Reliability of trial data			
Accuracy of trial data			
Participant's right, safety and wellbeing			
No impact on trial or participant			
Is the deviation an important deviation			
Yes			
No			
Number of participants with protocol deviation			

5.2 Withdrawal from treatment and/or follow-up

The numbers and percentages of participants who are lost to follow-up or withdraw will be reported by treatment allocation for each time point until the primary endpoint at 12 months post-randomisation. Reasons for withdrawal will also be summarised by treatment allocation in Table 4a and Table4b.

Table 4a: Summary of withdrawals and losses to follow from baseline to 1 year post-randomisation. For the purpose of this report, a participant is considered to be lost to follow-up at a particular visit if all follow-up outcome assessments following that visit are all missing.

	Operative fixation				Non-operative treatment			
	Withdrawals	%	Lost to follow-up	%	Withdrawals	%	Lost to follow-up	%
Totals								
Baseline								
4 weeks								
6 weeks								
3 months								
6 months								
1 year								

Table 4b: Summary of reasons for withdrawal by treatment allocation

Reason for Withdrawal	Operative fixation		Non-operative treatment		Total	
	n	%	n	%	N	%
Total Withdrawn						
Parent/guardian doesn't like the idea of being part of research						
Participant (child) doesn't like the idea of being part of research						
Parent/guardian doesn't want to complete questionnaires						
Participant (child) doesn't want to complete questionnaires						
No reason						
Other reason						

5.3 Baseline Characteristics

Baseline comparability of the randomised groups on stratification factors will be presented (Table 5). Baseline comparability of the randomised groups on other important prognostic factors and values of primary and secondary outcomes will be considered by treatment allocation. One of the stratification factors includes sites. A large number of sites is participating and a number of them has recruited a small number of participants. Sites with a small number of participants (<5) will be grouped together and marked as such when presenting baseline counts by site as this is how they will be handled in the analyses (Section 6.2.).

Variables will be presented by intervention arm numbers (with percentages) for binary and categorical variables and means and standard deviations (SDs), or medians and interquartile ranges (IQRs) for continuous variables will be presented; there will be no tests of statistical significance nor confidence intervals for differences between randomised groups on any baseline variable (Table 6-7).

Outcomes reported only for subgroup of participants will be clearly marked and footnotes explaining the nature of selective response e.g. DASH S/PA module is administered only to those participating at sports. Separate comparative Table presenting the descriptive statistics for the remaining of the outcomes at baseline for each subgroup with and without a measurement e.g. An additional baseline table of all outcomes measured for all for the group participating at sport and the group which does not.

Table 5: Stratification factors used in minimisation algorithm, by treatment allocation

Stratification Factor	Operative fixation		Non-operative treatment		Total	
	n	%	n	%	n	%
Centre						
Alder Hey Children's Hospital						
Bradford Royal Infirmary						
Bristol Royal Hospital for Children						
Countess of Chester Hospital						
Derriford hospital						
Hull Royal Infirmary						
James Cook University Hospital (Middlesbrough)						
Jenny Lind Children's Hospital (Norwich)						
Leeds general Infirmary						
Leicester Royal Infirmary						
Medway Maritime Hospital						
Nottingham University Hospital (Queen's Medical Centre)						
Royal Berkshire Hospital						
Royal Cornwall Hospital						
Royal Stoke University Hospital						
Salisbury District Hospital						
Sheffield Children Hospital						
St. Georges Hospital						
Sunderland Royal Hospital						
The Royal Belfast Hospital for Sick Children						
Tunbridge Wells Hospital						
University Hospital Coventry						
University Hospital Southampton						
University Hospital Wales (Cardiff)						
Whiston Hospital						
Dislocation status						
Not dislocated						
Dislocated						

Table 6: Baseline characteristics of participants and details of injury

Continuous	Operative fixation		Non- operative treatment		Total	
	mean	SD	mean	SD	mean	SD
Age						
Categorical	**n**	**%**	**n**	**%**	**n**	**%**
Gender						
Male						
Female						
Side of Injury						
Right elbow						
Left elbow						
Dominant Arm injury						
Yes						
No						
Unsure/Ambidextrous						
Mechanism of Injury						
Sporting injury						
Other						
Sport if mechanism of injury was sporting injury						
Dance						
Football						
Gymnastics						
Tennis						
Rugby						
Other						

Table 7: Baseline patient/proxy reported outcomes by treatment allocation

Questionnaire	Operative fixation			Non-operative treatment			Total		
	n	Mean	SD	n	Mean	SD	n	Mean	SD
PROMIS									
DASH S/PA									
Wong-Baker									
FACES									
EQ-5DY utility									
EQ-5DY VAS									

5.4 Unblinding

This is an unblinded study.

5.5 Treatment Compliance with Details of Intervention

The intervention groups in this trial are operative fixation or non-operative treatment of the medial epicondyle fracture of the elbow in children. As this is a one-off intervention, compliance is defined by the proportion of participants for each allocation that receive the treatment they were randomised to. The numbers and percentages of participants receiving each treatment are summarised, as well as those who did not receive the allocated treatment. Reasons are given in the cases where the participants did not receive the allocated treatment (Table 8).

Table 8: Compliance with allocated intervention

Treatment Received	Operative fixation		Non-operative treatment	
	n	%	n	%
Allocated Treatment				
Other				
Reasons for not receiving allocated treatment				
Clinical decision				
Child/Parent decision				
Lack of equipment				
Administrative error				
Other				

Details of treatment received are summarised in Table 9.

Table 9: Information about Intervention Received

	n	%
Operative Fixation:		
Screws		
Cannulated screws *Solid screws*		
Wires		
Buried *Unburied*		
Other		
Non-Surgical treatment:		
Above elbow full cast *Above elbow back slab*		
Other splint		
Bandage		
Other		
Post-Intervention Instructions:		
Cast off at X weeks Operative Fixation Non-Surgical treatment		
Immediate mobilisation as tolerated, with physiotherapy Operative Fixation Non-Surgical treatment		
Immediate mobilisation as tolerated, without physiotherapy Operative Fixation Non-Surgical treatment		
Rest until X weeks Operative Fixation Non-Surgical treatment		
Other Operative Fixation Non-Surgical treatment		

5.6 Reliability

To ensure consistency, validation checks of the data will be conducted. This will include checking for duplicate records, checking the range of variable values and validating potential outliers where possible (referring back to sites if necessary). As the data is collected electronically, many of these checks will be implemented automatically as part of the data entry procedure. Calculations and processes performed by a computer program, including the construction of derived data, will be checked by hand calculations. These checks will also confirm whether the data has been imported into the statistical software correctly and will check any merging of different datasets. Clarification will be sought from the trial office in the case of discrepancies.

For each variable, missing value codes will be checked for consistency and proportion of missing values per variable will be presented. Patterns of missing data will be explored. Where missing data imputation is used, imputed values will also be verified using the validation techniques described above.

6. ANALYSIS

The primary estimand of interest will be a treatment policy estimand (ITT population), using data from all randomised participants as randomised. This is chosen as the primary estimand to reflect the pragmatic nature of the study, and to identify the treatment effect regardless of any intercurrent events occurring. The elements of the primary estimand are summarised in Table 1.

The protocol also mandates a per-protocol analysis. The per-protocol analyses (section 4.2) will include all randomised participants who received their allocated treatment but excludes those randomised participants who were found ineligible post-randomisation, did not receive their allocated treatment.

Treatment switching as defined by NICE refers to the situation in a randomised controlled trial where patients switch from their randomly assigned treatment onto an alternative. Although it is widely recognised that ITT does not identify the true comparative effectiveness of the treatments under investigation there is disagreement on the acceptability of adjustment methods that can be used. Following NICE recommendations our ITT analyses is supplemented with further analyses in an attempt to adjust for switching [16] and described in section 6.4 sensitivity analyses..

The primary analysis and sensitivity analysis described in sections 6.2-6.4 focus on the primary estimand detailed in Table 1 which targets the average difference in PROMIS scores (ATE) between the two interventions in the combined groups randomised to receive each intervention. The resulting treatment effect represents the average difference in PROMIS scores we can expect if all participants in the target population were treated with non-operative treatment instead of being treated with operative fixation [17-19].

6.1 Outcome Definitions

Primary Outcome PROMIS Upper Extremity: measured pre-randomised (baseline), 4 weeks (routine follow-up) and then at 6 weeks, 3 months, 6 months and 12 months via electronic collection.

The primary outcome for this study is the functional recovery assessed using the Patient Report Outcomes Measurement Information System (PROMIS Bank v2.0: Upper Limb Score for Children Computer Adaptive Test (CAT) –PROMIS is a collection of patient-reported health status tools available for children and adults that were developed to be disease non-specific in collaboration with the USA National Institute for Health [20, 21]. This is a bounded score on a standardised scale (T-score) in the interval [10 – 58] (https://www.healthmeasures.net/images/PROMIS/manuals/Scoring_Manual_Only/PROMIS_Physical_Function_User_Manual_and_Scoring_Instructions_05Dec2023.pdf). These tools can be administered to healthy children, as well as to children with a variety of chronic health conditions. They are generally self-reported from the age of eight years, and proxy-reported below eight years. The PROMIS Paediatric item banks were developed using a strategic item generation methodology adopted by the PROMIS Network using item response theory. Field-testing occurred among 4,129 children aged eight to 17 years old [22]. All raw scores generated from PROMIS instruments are translated into standardized T-scores with a population mean of 50 and a standard deviation (SD) of 10. The population mean refers to the mean of the calibration sample, which, for paediatric and parent proxy instruments, is composed of a higher percentage of patients with chronic illness. Lower T scores indicate a worse outcome for upper limb function. PROMIS is available in full (30 questions), short-form (eight questions), or as a computer adaptive test 'CAT' (average eight questions). A CAT

enables the answer from one question to inform the choice of the next so that each child completing a CAT could answer a distinct set of questions to arrive at their score.

Secondary Outcomes

Sport/Performing Art Module of DASH: measured pre-randomised (baseline), 4 weeks (routine follow-up) and then at 6 weeks, 3 months, 6 months and 12 months via electronic collection.

This is a tool for recording details of sports and performing arts participation relating to upper limb function. Although not specifically developed in children, there was universal agreement, among children present at an 'Elbow Study Day' and members of the NIHR YPAG, that the language in this tool was appropriate for use among children who are able to comprehend other self-reported questionnaires used in this study [23].

Wong-Baker FACES Pain Rating Scale: measured pre-randomised (baseline), 4 weeks (routine follow-up) and then at 6 weeks, 3 months, 6 months and 12 months via electronic collection.

This is a validated self-reported ordinal assessment of pain [24] using a series of six facial expressions to illustrate pain intensity. A numerical rating is assigned to each face (from 0, "no hurt" to 10, "hurts worst"). It has been validated [25] for use among children from five years old. It is highly correlated to the visual analogue scale ($r = 0.90$, $p < 0.001$) and is widely used in clinical practice, forming part of the Royal College of Emergency Medicine 'Composite tool for the assessment of pain in children,'

(https://rcem.ac.uk/wp-content/uploads/2021/11/Pain_in_Children_2017_18_National_Report_Oct_2018.pdf (date last accessed 18 January 2024) and recommended in the NICE major trauma guidelines (https://www.nice.org.uk/guidance/ng39).

Quality of life (EQ-5D-Y): measured pre-randomised (baseline), 4 weeks (routine follow-up) and then at 6 weeks, 3 months, 6 months and 12 months via electronic collection.

This is the child-friendly version of the EQ-5D-3L, which has been adapted in terms of language for children aged eight to 11 years and for adolescents aged 12 to 18 years. A proxy version is available for younger children. Its age appropriateness in terms of feasibility, reliability, and validity in children and adolescents has been established [26, 27]. At the time of finalising this SAP, the adult value set is available and will be used. The HEAP will include supplementary analysis using available child value sets.

Complications: All complications are recorded at the 4-week routine follow-up and then at 6 weeks, 3 months, 6 months and 12 months via electronic collection.

Particular note will be made of complications related to the cast (e.g. pressure areas) or surgery (e.g. pain, wound infection, injury/irritation to the ulna nerve, implant irritation, screw cut-out, broken or retained metalwork, and the subsequent need to remove metal pins/screws), including hospital admission to manage these complications. These are summarised in the Table below.

Table 10: Comparison of complications & SAEs between treatment allocations

	Operative fixation		Non-operative treatment		OR (95% CI)		p-value
	n	%	n	%	Raw	Adj.	
Number of complications*		-		-	-	-	-
Participants with complications							
Wound problems (from surgery*)							
Elbow stiffness							
Unable to move fingers normally							
Patient Reported concerns							
Unable to feel fingers normally							
Ulna nerve problems (little & ring fingers) …							
Other							
Intra-operative complications		-		-	-	-	-
Participants with intra-operative complications							
Ulna nerve injury							
Fragmentation of bone							
Wire breakage							
Anaesthesia complications							
Post-operative bleeding							
Screw cut-out							
Other							
Immobilisation complications		-		-	-	-	-
Participants with immobilisation complications							
Skin complication owing to pressure area							
Discomfort							
Other							
Unplanned surgery/hospital admissions*		-		-	-	-	-
Participants with unplanned surgery/hospital admissions							
Fracture fixation/ revision fixation							
Removal of screw(s)/wire(s)							
Planned removal							
Irritation							
Other							
Nerve exploration (isolated procedure)							
Nerve exploration & screw/wire adjustment							
Nerve exploration and grafting							
Debride wound/ wound infection							
Microbiological confirmation of infection							
Antibiotics prescribed							
Other							

	Operative fixation		Non-operative treatment		OR (95% CI)		p-value
	n	%	n	%	Raw	Adj.	
Number of unforeseeable SAEs*		-		-	-	-	-
Participants with unforeseeable SAEs							

Radiographic outcomes and analysis will be considered in a separate document to this analysis plan.

Healthcare use: will be reported in the economic analysis and will include: missed school attendance at 6 weeks, 3 months, 6 months and 12 months and resource utilisation at 3 months, 6 months and 12 months [8]. The clinical report will only include comparative school attendance at 6 weeks, 3 months, 6 months and 12 months.

6.2 Analysis Methods

Analysis of Primary estimand (ATE): A Treatment Policy Analysis

Unadjusted summary statistics of the PROMIS scores will be displayed by treatment allocation using means and standard deviations. A multivariate linear mixed-effects regression model using repeated measures (level 1) nested within participants (level 2) will then be used to compare the treatment groups including all available data on randomised participants, at all-time points up to and including 12 months [28]. The model will include fixed effects to adjust for the minimisation factor used by the randomisation system: initial elbow dislocation status, and fixed effects for age and gender. A treatment by time interaction (as a categorical variable) will be included. The random part of the model will include random intercept terms for centres (level 3) and the individual participant (level 2) to account for the dependence of observations of participants within the same centre and observations within the same participant over time. At the individual participant level, it will also include a random coefficient for the effect of time (measurement occasion) and an unstructured covariance structure for the association between the random intercept and occasion as shown in the equations below describing the functional relationship between the outcome indexed by centre, individual participant and measurement occasions:

$$Y_{c,i,t} = \alpha_{c,i} + \beta_i Time_i + \theta\, Treatment_i + \gamma Treatment_i Time_i + \delta_k Covariate_{k,i} + \varepsilon_{i,t}$$

where c index centres, i individual participants and t measurement occasion (e.g. baseline, 1st, 2nd and 3rd follow-up visit) and $Time_i$ is a variable representing the time elapsed between randomisation and PROMIS questionnaire completed.

θ denotes the overall treatment effect and γ the interaction of treatment by time and covariates will be the stratification factors used in the randomisation protocol.

Note: Care needs to be taken to code observation occasions appropriately. Ideally, the exact timing of outcome measurement relative to the timing of randomisation will be used and if observation timings are highly unbalanced over time, time covariates will be centred around the mean time for fitting.

The random part of the model, partitions the variance to centre-level, individual participant and observation level and completes the model specification as follows:

Centre-level random effects

$$\alpha_{c,i} = \alpha_0 + u_{a,c} + u_{\alpha,i}$$

where $u_{a,c} \sim Normal(0, \sigma^2_{centre})$ quantifying between centre variation in baseline means and α_0 denoting the overall baseline mean

Individual participant-level random effects:

$$\beta_i = \beta_0 + u_{\beta,i}$$

$$\begin{pmatrix} u_{\alpha,i} \\ u_{\beta,i} \end{pmatrix} = MVN\begin{pmatrix} \sigma^2_{\alpha} & \sigma_{a,\beta} \\ \sigma_{a,\beta} & \sigma^2_{\beta} \end{pmatrix}$$

MVN denotes a multivariate normal distribution quantifying the between-individual variation in means and linear change over time, respectively along with their correlation and $\varepsilon_{i,t} \sim Normal(0, \sigma^2_{i,t})$ the residual observational level variance.

This will be compared with a model assuming independence between the random intercept and slope with respect to reductions in the likelihood function and the precision of the estimate for the random slope.

Non-linearity for the effect terms will also be considered using restricted cubic splines in the fixed part of the model. Even when highly non-linear patterns are evident, the limited number of measurement occasions—only five, including the baseline—restricts the complexity needed for an adequate fit. This means that we will need at most 2-3 knots, which should be positioned at the centiles of the time axis distribution [29] and choosing the model with the smallest deviance and the simplest functional form for the effect of time in the fixed part of the model, will be used after considering a maximum of three models: (a) linear (b) cubic spline with 2 knots (c) cubic spline 3 knots.

Sites with less than 5 participants will be grouped together to allow reliable estimation [30, 31].

The adjusted mean differences (treatment effect) will be presented, along with 95% confidence intervals and p-values.

As a supporting analysis to check model assumptions, a simple analysis of covariance (ANCOVA) of the primary outcome at 12 months will also be presented, adjusting only for the baseline PROMIS score, following protocol analysis specifications.

Model Checking and normalising transformations: Assumptions of normality will be assessed graphically looking at residual and quantile-quantile (QQ) plots, and data transformation will be considered if model assumptions (particularly normality of residuals) are clearly violated. Mixed-effects models are robust to small deviations from normality in the residuals; however, if approximate normality cannot be achieved the data normalising transformations will be considered and then proceed fitted the same mixed effects linear model described above on the transformed scale. We will use the logistic transformation which is a transformation recommended for bounded outcomes [32]. Data will be first transformed to a (0, 1) scale (transformed y_new=(y-min(y))/(max(y)-min(y)), where y represents the outcome), following Smithson et al [33] and using the logistic transformation, in line with the bounded nature of the measurement scale. Treatment effect estimates will be presented on the original scale by post-estimation back transforming.

Examination of residuals following model fitting will be undertaken including exploration of dependencies with observed covariates. Consideration will be given to augmenting the model to heteroscedastic (level 1) residuals especially when the model is fitted in the original scale, where strong patterns are observed between residual errors and covariates and variance increases with the mean [34]. It is expected that normalising transformation using the logistic transform will have a variance stabilising effect on the fitted model and

maybe preferable as it would result to a simplified model. Comparison of the log-likelihood values of a small number of competing models will be used to decide whether model (higher log-likelihood values indicate a better fitting model).

Outcomes measured in certain subgroups only: The secondary outcome DASH Sports/Performing arts module is only completed if the participant indicates regular participation in sport or musical activities. The analysis of this secondary outcome, **targeting the average treatment effect in the subgroup completing the DASH Sports/Performing arts module,** will use the same methods described for the primary outcome but augmented with propensity score weighting. As this subgroup is exclusively observed among participants engaged in sports and musical activities, this could probably lead to imbalances in covariates between the comparison groups. The propensity weighting scheme is selected because it maximises the efficiency of the estimate and provides an unbiased estimate of the subgroup average treatment effect due to randomisation. [35, 36]. The inverse probability treatment weighting (IPTW) scheme for the propensity of receiving the randomised treatment will be modelled as a function of baseline covariates including subgroup membership, age, sex, stratification factors, initial dislocation status, mechanism of injury, baseline pain score. All interactions between group membership and the rest of the baseline covariates will be included in this adjustment set [37].

6.3 Missing Data

Missing data will be minimised by careful data management. Missing data will be described with reasons given where available; the number and percentage of individuals in the missing category will be presented by treatment arm. All data collected on data collection forms will be used, since only essential data items will be collected. No data will be considered spurious in the analysis since all data will be checked and cleaned before analysis.

Missing covariate data: If the amount of missing covariate data leads to exclusion of less than 5% of participants, and it is implausible that such exclusion is associated with outcomes or is not associated with observed outcomes, then missing data will be considered ignorable. Fully adjusted analyses will be based on all randomised using imputed covariate data using single conditional imputation. Sensitivity analyses will include adjusted analyses based on those with complete covariate profile.

If the amount of missing covariate data leads to more than 5% and auxiliary variables can be identified that are associated with missingness with predictive value for the missing data then multiple imputation will be employed, including observed outcomes, randomised treatment, auxiliary covariate data predictive of missingness and all variables which will be used in adjusted analyses. The imputation model will include all participants randomised in the group originally assigned.

Missing outcome data: The primary analysis method proposed is reasonably robust to missing at random (MAR) data [38] conditional on all outcome measurements on all other visits and other covariates included in the model (Table 1).

Drop-out indicators will be created and we will report the counts of participants dropping out by arm, as well as the follow-up visit they dropped out (Table 4a). For the purpose of our main analyses reporting outcomes up to 12 months, we consider a participant as having drop-out at a particular visit if all follow-up outcome assessments following that visit are all missing.

We will then examine any association of drop-out with important prognostic factors such as age, sex, nerve injury at presentation, BMI, diabetic status, smoking status, concomitant injuries which affect limb function, clinically significant SAEs relating to the interventions (Section 6.7), and baseline PROMIS scores as well as follow-up PROMIS scores observed before drop out. If significant associations are found with pre-drop out follow-up scores then we will fit an informative missingness model following Diggle & Kenward [39] (e.g. using xteregress and/or GLLAMM in Stata) where the longitudinal PROMIS scores will be modelled jointly with the

drop-out indicator. This analysis will be undertaken only if there is evidence of dependence of drop-out indicators on observed outcome data.

6.4 Sensitivity Analysis

Primary estimand: Average Treatment Effect (ATE) under a treatment policy strategy

Sensitivity analyses will be undertaken for the primary estimand [16] where all randomised patients will be analysed according to their treatment allocation and adjusting for treatment switching using **Propensity Score (Quantile) Regression adjustment.** Treatment switching post-randomisation due to individual patient and treatment characteristics prognostic of the outcome, can affect both the treatment effect and its precision, especially when the related covariate profile is highly unbalanced between the two groups. Inclusion of the propensity score of treatment received in the analysis model can adjust for such imbalances yielding more accurate effect estimates under the assumption that we have measured all confounding variables and that each participant has a nonzero probability to be treated with surgical fixation (positivity assumption) [18, 19, 40] .

The initial step involves fitting the propensity score model. The propensity score model is a logistic regression model, modelling the probability of treatment received as a function of covariates that are prognostic for the outcome and individual participant characteristics associated with treatment receipt. The estimated Propensity Score (PS) derived from this model is an approximation of the true PS. Overfitting is not a problem in PS estimation as long as due diligence is applied to select covariates that are the most likely risk factors for the outcomes. Then this is used by each approach in different ways.

In the Regression adjustment approach, we include the quantiles of the propensity score in the analysis model as covariates and including treatment interactions with the PS-quantiles. This will correct imbalance and can reduce the standard error of the effect estimates [41, 42]. The balancing properties depend on the amount of overlap of the propensity score distribution in the two groups with associated bias-variance trade-offs whereby the more closely matched the comparison groups the less biased the treatment effect at the cost of increased variance.

The distribution of the PS for each randomised group will be examined graphically. IPTW balance will also be examined. The balance of the confounder after weighting the contributions of individuals (baseline and treatment characteristics used to model the PS) between those who assigned to surgical fixation and those assigned to functional bracing will be documented as standardised differences before and after weighting for each confounder used (e.g. using Stata's rebalance command). Although, there is no definitive value at which the treatment is considered unbalanced, a variance ratio less than 0.5 indicates that the data is not balanced and indicates the PS model may need to be revisited and/or trimming or removing data at the extreme of the distribution of the weights (e.g. the 5^{th} and 95^{th} percentiles) maybe needed.

The above estimator targets the average difference in PROMIS scores between the two groups (ATE). **The resulting treatment effect represents the average difference in PROMIS scores we can expect if all participants in the target population were treated with non-operative treatment fixation instead of being treated with operative fixation** [17].

Missing outcome data: The primary analysis method proposed is reasonably robust to missing at random (MAR) data [38]. Treatment effects will be compared to estimates resulting from the same model used in the primary (ITT) analyses after further adjusting for additional factors that are both prognostic of the outcome and associated with drop-out.

Informative drop out: Sensitivity analysis will also be undertaken on the primary analysis by imputing missing data under different missing not at random (MNAR) assumptions for the model targeting the ATE estimand and using the ITT population. This will be achieved using δ-based imputation following Cro et al. [43], where our primary is fitted and adding an offset δ-term with values set to the expected value of the missing data. This is done to assess the impact of unobserved participants having a worse or better response than those observed (e.g. implemented through the rctmiss Stata command) and is consistent with an ITT primary analysis. We will vary the values of the offset term using a range of different means (+/- 1 SD of PROMIS scores).

6.5 Pre-specified Subgroup Analysis

We will investigate treatment effects across the following subgroups of clinical interest as follows: key sub-groups age-group (8+ or younger) and initial dislocation status and initial duration of immobilisation. The purpose of subgroup analyses is to investigate if the estimated treatment effects are relatively consistent across subgroup and for this extent will be viewed as exploratory. None of the included subgroups are based on post-randomisation patient characteristics or events [15].

If there are too few participants in any subgroup (<=15) or one treatment arm of a subgroup (<=5), the analysis for that subgroup will not be conducted. Consistency of effects will be undertaken on the ITT population for the primary estimand. Treatment effects for subgroups will be derived by introducing an interaction term of treatment with the covariate representing the clinical grouping of interest in the main model for the primary estimand. Interactions will be examined in separate models for each clinical grouping of interest. If the clinical grouping is measured in a way that gives rise to a continuous variable, the interaction term will include the variable on the continuous scale for testing treatment effect modification by the characteristic of interest. Treatment differences will be then summarised and reported for changes over high-density areas of the distribution of such covariates i.e. 25%, 50% and the 75% centiles. Age groupings will be used here as these age groupings also represent parental help when responding to PROMIS self-report. The estimated treatment difference and confidence intervals will be reported and presented in forest plots.

6.6 Supplementary/ Additional Analyses and Outcomes

The SCIENCE study will follow up with participants annually up until 16 years of age. This will be to draw inference on long-term pain and function, as well as looking at additional surgery. A separate appendix to the SAP will document the planned long-term follow-up analyses. Analysis plans for radiographic outcomes will be documented separately.

6.7 Harms

This is a low risk, pragmatic trial where both of the interventions are in common use. Thus, we do not anticipate many SAEs. All adverse events are submitted to the SCIENCE central office ONLY if they arise 12 months post-randomisation (i.e. a period deemed related to the intervention) and they fall under the category of an SAE: events resulting in Death; Life-threatening; required hospitalisation; prolonged hospitalisation; congenital abnormality of birth defect; other important medical event which may require medical or surgical intervention to prevent one of the serious outcomes listed. SAEs potentially related and unexpected are recorded on the trial's database and assessed for seriousness, causality and expectedness (Table 11).

Table 11: List of serious adverse events

Allocation	Diagnosis	Timing of onset of event[1]	Description of event (including signs & symptoms)	Action taken to deal with event (including any treatment)	Reason for seriousness[2]	Causality: is the event related to the intervention?[3]	Expectedness (in relation to what is known about the intervention)[4]

[1]: post-randomisation weeks
[2]: *Death; Life-threatening; required hospitalisation; prolonged hospitalisation; congenital abnormality of birth defect; other important medical event which may require medical or surgical intervention to prevent one of the serious outcomes listed.*
[3]: *Possibly related; Probably related; Definitely related*
[4]: *Expected; Unexpected*

Only unexpected SAEs potentially related to the intervention. Adverse events that are foreseeable and are not SAEs are recorded in the Complications section and are reported as outcomes.

6.8 Health Economics and Cost Effectiveness (where applicable)

The statistician is not undertaking this analysis. A separate Health Economics Analysis Plan (HEAP) will be written by the trial health economist and all cost effectiveness analysis will be undertaken following that plan by the health economist.

7. VALIDATION OF THE PRIMARY ANALYSIS

To validate the primary outcome and key secondary outcomes a statistician not involved in the trial will independently repeat the analyses detailed in this SAP, by using different statistical software (if possible). The results will be compared, and any unresolved discrepancies will be reported in the Statistical report (See OCTRU SOP STATS-005 Statistical Report). If necessary, this will include derivation of the primary and key secondary outcomes from raw data.

Validation will be undertaken for the main model of the primary analysis only and will include: the primary outcome for analysis is PROMIS at all time points up to 12 months, and key secondary outcomes: DASH S/PA at all time points up to 12 months and EQ5D-5DY at all time points up to 12 months.

8. SPECIFICATION OF STATISTICAL PACKAGES

All analysis will be carried out using appropriate validated statistical software such as STATA, SAS, or R. The relevant package(s) and version number(s) will be recorded in the Statistical report.

9. PUBLICATION

This study will be/has been conducted as part of the portfolio of trials in the registered UKCRC Oxford Clinical Trials Research Unit (OCTRU) at the University of Oxford. It will follow/has followed their Standard Operating Procedures ensuring compliance with the principles of Good Clinical Practice and the Declaration of Helsinki and any applicable regulatory requirements.

10. REFERENCES

1. *ICH E9 (R1):Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials* 2020.
2. Gamble, C., et al., *Guidelines for the Content of Statistical Analysis Plans in Clinical TrialsGuidelines for the Content of Statistical Analysis Plans in Clinical TrialsGuidelines for the Content of Statistical Analysis Plans in Clinical Trials.* JAMA, 2017. **318**(23): p. 2337-2343.
3. Pathy, R. and E.R. Dodwell, *Medial epicondyle fractures in children.* Current Opinion in Pediatrics, 2015. **27**(1): p. 58-66.
4. Kamath, A.F., et al., *Operative versus non-operative management of pediatric medial epicondyle fractures: A systematic review.* Journal of Children's Orthopaedics, 2009. **3**(5): p. 345-357.
5. Knapik, D.M., et al., *Outcomes of Nonoperative Pediatric Medial Humeral Epicondyle Fractures With and Without Associated Elbow Dislocation.* Journal of Pediatric Orthopaedics, 2017. **37**(4): p. e224-e228.
6. Marson, B.A., et al., *Interventions for treating supracondylar elbow fractures in children.* Cochrane Database of Systematic Reviews, 2022(6).
7. Mautner, B.K. and J. Blazuk, *Overuse Throwing Injuries in Skeletally Immature Athletes — Diagnosis, Treatment, and Prevention.* Current Sports Medicine Reports, 2015. **14**(3): p. 209-214.
8. Achten, J., et al., *Protocol for Surgery or Cast of the EpicoNdyle in Children's Elbows (SCIENCE).* Bone & Joint Open, 2024. **5**(1): p. 69-77.
9. Ferguson, D. and D.C. Perry, *Managing displaced fractures of the medial humeral epicondyle in children.* The Bone & Joint Journal, 2024. **106-B**(3): p. 224-226.
10. Kang, M., et al., *Incorporating estimands into clinical trial statistical analysis plans.* Clinical Trials, 2022. **19**(3): p. 285-291.
11. Morgan, E., et al., *Establishing clinical meaning and defining important differences for Patient-Reported Outcomes Measurement Information System (PROMIS ®) measures in juvenile idiopathic arthritis using standard setting with patients, parents, and providers.* An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation - Official Journal of the International Society of Quality of Life Research, 2017. **26**(3): p. 565-586.
12. Thissen, D., et al., *Estimating minimally important difference (MID) in PROMIS pediatric measures using the scale-judgment method.* An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation - Official Journal of the International Society of Quality of Life Research, 2016. **25**(1): p. 13-23.
13. Group, D.S., *A proposed charter for clinical trial data monitoring committees: helping them to do their job well.* The Lancet, 2005. **365**(9460): p. 711-722.
14. Cooperation, S., *Stata 15.* Stata Cooperation, College Station, TX, 2017.
15. Yusuf, S., et al., *Analysis and Interpretation of Treatment Effects in Subgroups of Patients in Randomized Clinical Trials.* JAMA, 1991. **266**(1): p. 93-98.
16. Latimer, N.R., et al., *TREATMENT SWITCHING: STATISTICAL AND DECISION-MAKING CHALLENGES AND APPROACHES.* International Journal of Technology Assessment in Health Care, 2016. **32**(3): p. 160-166.
17. Cole, S.R. and M.A. Hernán, *Constructing Inverse Probability Weights for Marginal Structural Models.* American Journal of Epidemiology, 2008. **168**(6): p. 656-664.
18. Efron, B. and D. Feldman, *Compliance as an Explanatory Variable in Clinical Trials.* Journal of the American Statistical Association, 1991. **86**(413): p. 9-17.
19. Joffe, M.M., T.R.T. Have, and C. Brensinger, *The compliance score as a regressor in randomized trials.* Biostatistics, 2003. **4**(3): p. 327-340.
20. Cella, D., et al., *The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008.* Journal of Clinical Epidemiology, 2010. **63**(11): p. 1179-1194.
21. Morgan, E.M., et al., *Establishing clinical meaning and defining important differences for Patient-Reported Outcomes Measurement Information System (PROMIS®) measures in juvenile idiopathic arthritis using standard setting with patients, parents, and providers.* Quality of Life Research, 2017. **26**(3): p. 565-586.
22. Irwin, D.E., et al., *Sampling plan and patient characteristics of the PROMIS pediatrics large-scale survey.* Quality of Life Research, 2010. **19**(4): p. 585-594.
23. Hudak, P.L., et al., *Development of an upper extremity outcome measure: The DASH (disabilities of the arm, shoulder, and head).* American Journal of Industrial Medicine, 1996. **29**(6): p. 602-608.
24. Wong, D. and B. CM, *Pain in children: comparison of assessment scales.* Pediatr Nurs, 1988. **14**(1): p. 9-17.
25. Tomlinson, D., et al., *A Systematic Review of Faces Scales for the Self-report of Pain Intensity in Children.* Pediatrics, 2010. **126**(5): p. e1168-e1198.
26. Eidt-Koch, D., T. Mittendorf, and W. Greiner, *Cross-sectional validity of the EQ-5D-Y as a generic health outcome instrument in children and adolescents with cystic fibrosis in Germany.* BMC Pediatrics, 2009. **9**(1): p. 55.

27. Wille, N., et al., *Development of the EQ-5D-Y: a child-friendly version of the EQ-5D.* Quality of Life Research, 2010. **19**(6): p. 875-886.
28. Goldstein, H., M.J.R. Healy, and J. Rasbash, *Multilevel time series models with applications to repeated measures data.* Statistics in Medicine, 1994. **13**(16): p. 1643-1655.
29. Harrell, F.E., *Multivariable Modeling Strategies*, in *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*, F.E. Harrell, Editor. 2001, Springer New York: New York, NY. p. 53-85.
30. Clarke, P., *When can group level clustering be ignored? Multilevel models versus single-level models with sparse data.* Journal of Epidemiology and Community Health, 2008. **62**(8): p. 752.
31. Theall, K.P., et al., *Impact of small group size on neighbourhood influences in multilevel models.* Journal of Epidemiology and Community Health, 2011. **65**(8): p. 688.
32. Lesaffre, E., D. Rizopoulos, and R. Tsonaka, *The logistic transform for bounded outcome scores.* Biostatistics, 2006. **8**(1): p. 72-85.
33. Smithson, M. and J. Verkuilen, *A Better Lemon Squeezer? Maximum-Likelihood Regression With Beta-Distributed Dependent Variables.* Psychological Methods, 2006. **11**: p. 54-71.
34. Goldstein, H., *Heteroscedasticity and Complex Variation*, in *Encyclopedia of Statistics in Behavioral Science*. 2005.
35. Williamson, E.J., A. Forbes, and I.R. White, *Variance reduction in randomised trials by inverse probability weighting using the propensity score.* Statistics in Medicine, 2014. **33**(5): p. 721-737.
36. Zeng, S., et al., *Propensity score weighting for covariate adjustment in randomized clinical trials.* Statistics in Medicine, 2021. **40**(4): p. 842-858.
37. Lin, W., *Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique.* The Annals of Applied Statistics, 2013. **7**(1): p. 295-318, 24.
38. Sullivan, T.R., et al., *Should multiple imputation be the method of choice for handling missing data in randomized trials?* Statistical methods in medical research, 2018. **27**(9): p. 2610-2626.
39. Diggle, P. and M.G. Kenward, *Informative Drop-Out in Longitudinal Data Analysis.* Journal of the Royal Statistical Society. Series C (Applied Statistics), 1994. **43**(1): p. 49-93.
40. White, I.R., *Uses and limitations of randomization-based efficacy estimators.* Statistical Methods in Medical Research, 2005. **14**(4): p. 327-347.
41. D'Agostino Jr., R.B., *Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group.* Statistics in Medicine, 1998. **17**(19): p. 2265-2281.
42. Rubin, D.B. and N. Thomas, *Matching Using Estimated Propensity Scores: Relating Theory to Practice.* Biometrics, 1996. **52**(1): p. 249-264.
43. Cro, S., et al., *Sensitivity analysis for clinical trials with missing continuous outcome data using controlled multiple imputation: A practical guide.* Statistics in Medicine, 2020. **39**(21): p. 2815-2842.

APPENDIX: GLOSSARY OF ABBREVIATIONS

AE	Adverse Event
BNF	British National Formulary
CAT	Computer Adaptive Test
CHI	Community Health Index number (Scotland)
CI	Chief Investigator
CRF	Case Report Form
DSMC	Data and Safety Monitoring Committee
EQ-5D-Y	EuroQol – Youth
GCP	Good Clinical Practice
HE	Health Economy/Economist
HEAP	Health Economic Analysis Plan

HTA	Health Technology Assessment
HRA	Health Research Authority
ISAP	Interim Statistical Analysis Plan
MAR	Missing at Random
MCID	Minimally Clinically Important Distance
NICE	National Institute for Health and Care Excellence
OCTRU	Oxford Clinical Trials Research Unit
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROMIS	Patient Reported Outcome Measurement System
QA	Quality Assurance
QALY	Quality Adjusted Life Year
RCT	Randomised Controlled Trial
REC	Research Ethics Committee
SAP	Statistical Analysis Plan
SAR	Serious Adverse Reaction
SUSAR	Serious Unexpected Adverse Reaction
TMG	Trial Management Group
TSC	Trial Steering Committee
YPAG	Young Persons Advisory Group