HUSH

The HUmeral SHaft fracture trial: A multi-centre prospective randomised superiority trial of surgical versus non-surgical interventions for humeral shaft fractures in patients aged 18 years or older

Statistical Analysis Plan

Version 3.0 – 16May2025

Linked to SAP - Data definitions and Tables
Version 2.1 – 28Oct2024

Based on Protocol version 6.0 – 02Nov2022

Trial registration: ISRCTN 17108318

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

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 for Health Research funded multi-centre prospective randomised superiority trial of surgical versus non-surgical interventions for humeral shaft fractures in patients aged 18 years or older (HUSH).** 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.

Integral to this Statistical Analysis Plan (SAP) is the SAP – Data Definitions and Tables document which will include full detailed descriptions of all key outcomes, including their definition, generation and how they will be reported at the end of the study. These two documents should be read in tandem.

1.1 Key personnel

List of key people involved in the drafting and reviewing this SAP, together with their role in the trial and their contact details.

Author(s) (Trial statistician):

Reviewers:

Approver:

1.2 Changes from previous version of SAP

This is the third version of the SAP. The table below gives a summary of changes from the previous versions of SAP.

Version number Issue date	Protocol Version & Issue date	Significant changes from previous version together with reasons
V1.0_22Oct2024	V6.0 02Nov2022	Not applicable as this is the 1st issue
V2.0_28Oct2024	V6.0 02Nov2022	Minor changes to reference updated SAP DT&T where clarifications are added for adjusted analyses and also include Table with SAE listings. The CONSORT flowchart is also updated to include follow-up points
V3.0_16May2025	V6.0 02Nov2022	Minor changes to correct typographical errors in Section 6.2 and Section 6.6. There was no

		impact on any analysis as a result of these typographical errors.

2. BACKGROUND AND OBJECTIVES

This pragmatic superiority randomised controlled trial aims to evaluate the clinical and cost-effectiveness of functional bracing, compared to surgical fixation for treating humeral shaft fractures in patients over the age of 18 years [3].

The primary objective is to quantify and draw inferences on observed differences in function using the Disabilities of Arms Shoulders and Hand patient-reported outcome questionnaire (DASH) between functional bracing and surgical fixation at 12 months. Secondary objectives include quantifying observed differences and drawing inferences in patient-reported outcomes, pain and recovery profile, risk of complications, resource use and comparative cost-effectiveness and duration of time off working for participants of working age, between the intervention group at 12 months.

The estimand for the primary objective and the analysis of the primary outcome is described in Table 1.

Table 1: Estimand-to-analysis table template

Primary Objective: *Evaluate* the superiority surgical fixation against functional bracing compared to at 12 months for the treatment of humeral shaft fractures in patients over the age of 18 years.	
Estimand: *D*ifference in function scores between treatment groups, at 12 months post-randomisation, in patients aged 18 years or older with humeral shaft fractures for whom functional bracing is a clinically appropriate option according to current clinical practice.	
Treatment: *Surgical fixation vs Functional bracing (current standard care)*	
Estimand	***Analysis***
Target population *Adults aged 18 years or older with humeral shaft fractures* for whom functional bracing is a clinically appropriate option according to current clinical practice.	***Analysis set*** *All randomised participants.*
Variable Function measured on the Disabilities of Arms Shoulders and Hand patient-reported outcome scale (DASH)	***Outcome measure*** *Continuous outcome bounded between 0 (no disability - to 100 (severe disability) scale, at 12 months post-randomisation*
Handling of intercurrent events Participants allocated to surgical fixation who receive functional bracing at the point of initial treatment or within six months post-randomisation Participants allocated to functional bracing who received surgical fixation Any additional interventions received in either arm Will ALL be handled under a treatment-policy	***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 stratification 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.

strategy[1] for handling them and where intercurrent events are considered part of the treatment being compared.	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 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 DASH score). This is done to assess the impact of drop-out on inferences

Population-level summary measure	***Analysis approach***
Average difference in scores between the groups at 12 months if all eligible patients were treated with surgical fixations instead of functional bracing	Analysis of repeated measures of DASH 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 mixed effects model. Sensitivity analyses: Primary estimand (ITT) for the average treatment effect (ATE) using (a) inverse probability weighting (IPTW) based on the propensity score for receiving the allocated treatment and using the stabilised version of the IPTW estimator and (b) 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, 4]*

3. STUDY METHODS

3.1 Trial Design/framework

This trial is a pragmatic, multi-centre, two-arm, parallel group, randomised controlled superiority clinical trial with parallel economic analysis and direct patient follow-up to 1-year. The trial will employ 1:1 treatment allocation, stratified by centre, age, and nerve injury with patients randomised to either functional bracing or surgical fixation, based on the surgeons' usual surgical practice.

The trial is split into two phases – a pilot phase and a main phase. Study procedures during both phases will be identical as per this protocol. During the initial pilot phase of 6 months, 8 centres will be opened to recruitment. At the end of the pilot phase, there a one decision point with regards to the continuation of the trial. The stop-go criteria are given in protocol **Table A**. If recruitment fails to reach 20 participants by the end of the pilot phase (six months after trial opening), the DSMC may recommend that the trial is terminated.

Screening and subsequent recruitment for the main phase will occur at a minimum of 16 NHS hospitals over a 21-month period. All treatments are standard NHS treatments and will be conducted at the recruiting centres. Participants will be followed up clinically as per standard hospital policy. They will be followed-up via postal or electronic questionnaires by the central trial team for a period of 12 months.

Table A: Stop-Go criteria for the main trial

	Actual recruitment at the end of the pilot phase (6 Months)		
% Threshold	≤67%	68-99%	100%
Recruitment rate (per centre per month)	≤0.5	0.6-0.7	>0.8
Number of sites opened	<8	8	>8
Trial Recruitment	<20 participants	21-29 participants	>30 participants
Stop-go outcome	Recruitment not feasible; decision not to proceed	Review recruitment strategies. Report to TSC. Continue but modify & monitor closely	Recruitment feasible; proceed with study

3.2 Randomisation and Blinding

Once informed consent has been given, the participant will be randomised by the local research team using a centralised web- service. The randomisation will be a 1:1 basis, using a validated computer randomisation programme managed through a secure (encrypted) web-based service by the Oxford Clinical Trials Research Unit (OCTRU), with a minimisation algorithm to ensure balanced allocation across treatment groups, stratified by centre, age (< 50 years vs ≥ 50 years), nerve injury at presentation (Yes/No). The minimisation algorithm will include a probabilistic element and a small number of participants randomised by simple randomisation at the start of the trial to seed the algorithm to ensure unpredictability of treatment allocation.

On randomisation of a participant the central office, main site contact and local study team will be notified. This will take place via an automated email as part of the randomisation process.

A paper-based randomisation system will be in place for use in emergencies, e.g. if the web-based randomisation service is not functioning, an event that is rare with this service.

Due to the nature of the trial, neither the treating team nor the participants will be blinded. The local research team reviewing hospital records will also not be blind to the treatment allocation. Any radiographs collected will be reviewed by an independent adjudication committee who, due to the presence of the metalwork, will also not be able to perform their assessments blinded.

Full details of the randomisation and emergency randomisation are available in **HUSH_RBP_V1.0_14Sep2020**, stored in the confidential statistical section of the TMF.

3.3 Sample Size

The minimum clinically important difference for the DASH questionnaire has been identified as 10 points, and the standard deviation available from the literature is variable with the closest to our target population being 21.7 [5]. A standardized effect size of 0.4 (a small to moderate effect size) equates to a difference of 10 points when the standard deviation is as high as 25 or a difference of 8 points when it is as low as 20. At 90% power

and 5% (2-sided) significance, the proposed sample size needed is 266 (133 per treatment arm) participants providing data at 12 months in order to detect a standardised effect size of 0.4. Allowing for 20% loss to follow-up yields an overall target of 334 (167 per arm). These calculations are based on the primary outcome of DASH at 12 months.

In summary, a minimum of 266 participants with primary outcome data (DASH at 12 months) will provide a definitive answer to the research question with 90% power and 5% (2-sided) significance to detect a standardised effect size of 0.4. To achieve this number and allowing for loss to follow-up we aim to randomise 334 participants.

3.4 Statistical Interim Analysis, Data Review and Stopping guidelines

A Data and Safety Monitoring Committee (DSMC) will oversee the conduct of the trial and safety of participants by reviewing accruing blinded data during recruitment and follow-up including any planned interim analyses. The DSMC follows the charter **HUSH_DSMC Charter_V1.0_05May2020** stored in the TMF. The DSMC will review accruing data and summaries of the data presented by treatment group. They will assess the screening algorithm against the eligibility criteria. They will also consider emerging evidence from other related trials or research and review related SAEs that have been reported.

Full details of the data that will be shown to the DSMC are available in the DSMC report template, **HUSH_DSMCTemplate_V1.0_26Apr2021** stored in the confidential section of the TMF. The DSMC will review the sample size assumptions approximately half-way through recruitment to the study to ensure that this sample size would be able to provide a definitive answer to the research questions.

This trial will have one decision point, at the end of the pilot phase. The DSMC will advise the TSC on continuation of the trial at the end of the pilot phase. Details are described in section 3.1.

There is no interim analysis planned for this trial.

3.5 Timing of Analysis

The final analysis will be conducted once the final participants has reached their 12 months follow up time period, and the data has been received, cleaned and finalised.

3.6 Blinded analysis

A “blinded” review and cleaning of data (not separated by intervention groups will be undertaken prior to the final data lock. However, such review or data cleaning cannot be undertaken in a “blinded” way when data summaries, or participant specific information requires clarification related to receipt of intervention. Data distributions will be summarised, a look at missing data patterns will be done, and data quality checks will be performed. The identification of participants who will be excluded from the per-protocol analysis will also be finalised as far as possible. The final analysis will be unblinded. The trial statistician and study staff are not blinded to the allocation at any point, but study staff at NHS centres do not have access to any data from other study centres.

3.7 Statistical Analysis Outline

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. Standard statistical summaries and graphical plots will be presented for the primary outcome measure and all secondary outcome measures.

The analysis for the trial will be based on the intention-to-treat (ITT) population. The DASH at 12 months is the primary outcome and it will be analysed using a mixed-effect linear regression model. The DASH at all time points (except the DASH at baseline) will be the dependent variable and the independent variable will be the

interaction between treatment group and follow-up time points. The stratification factors for randomisation (recruitment centre, age and nerve injury at presentation) and baseline (preinjury) DASH score will be adjusted in the model. The random effects of the model will be recruitment centre to account for the heterogeneity that may arise due to variation between centres. The other variables will be incorporated as fixed effects. The treatment effect will be the estimated difference at 12 months, it will be reported alongside 95% confidence interval and will be used to determine superiority.

A fully adjusted model and an unadjusted model will also be run. For the fully adjusted mixed-effect model, in addition to the factors that have already been adjusted in the main model, other important prognostic factors (sex, diabetic-status, smoking status, Body Mass Index (BMI) and concomitant injuries which affect limb function and additional surgery) will also be included using the same methods.

Subgroups based on type of surgery/brace and stratification factors, and other baseline participant characteristics including, age, sex, nerve injury at presentation, fracture position (middle 1/3 fracture, proximal 1/3 fracture and distal 1/3 fracture) and fracture type (simple or complex fracture), mode of injury and BMI will be explored using treatment by subgroup interactions [6]. 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.

The analysis will be repeated using the per-protocol (PP) population for sensitivity analyses. If a substantial amount of non-compliance is observed or if the non-compliance is selective, then a complier average causal effect (CACE) analyses will be undertaken [7] with details in section 6.6.

4. STATISTICAL PRINCIPLES

4.1 Statistical Significance and Multiple Testing

There is no multiple testing as only a single primary outcome is considered. Therefore, the significance level used will be 0.05 and 95% confidence intervals will be 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.

No interim analyses of any primary or secondary outcomes were planned or performed during the study.

4.2 Definition of Analysis Populations

The intention-to-treat (ITT) population will include all participants with available data at time-points up to and including 12 months in the randomised groups to which they were allocated, regardless of the treatment they actually received.

The per-protocol (PP) population will include all participants who received their allocated treatments and did not have any major protocol deviations.

Protocol deviation criteria:

1. Participants recruited to trial who did not fully satisfy the eligibility criteria for the study.
2. Participant who did not receive the allocated treatment.
3. Randomised in error for any reason other than eligibility.

The per-protocol population is relevant to a secondary estimand detailed in section 6 under a principal stratum strategy and which targets the average treated effect for those treated with surgical fixation in supplementary analyses. Major protocol deviations and the per-protocol population will be finalised following a blinded review of the data prior to the final data lock.

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

Figure 1 shows the flow of participants through each stage of the trial, including numbers of participants randomly assigned, receiving intended treatment, completing the study protocol, and analysed for the primary outcome is provided following the updated CONOSRT guideline using the extensions for non-pharmacological treatment interventions and patient reported outcomes [8]. Protocol violations/deviations and information relating to the screening data including the number of ineligible patients randomised, together with reasons. Information on number of participants screened, found to be ineligible (with reasons available in Table 1), refused to participate (with reasons available in Table 2) will also be included.

Figure 1 The CONSORT flowchart

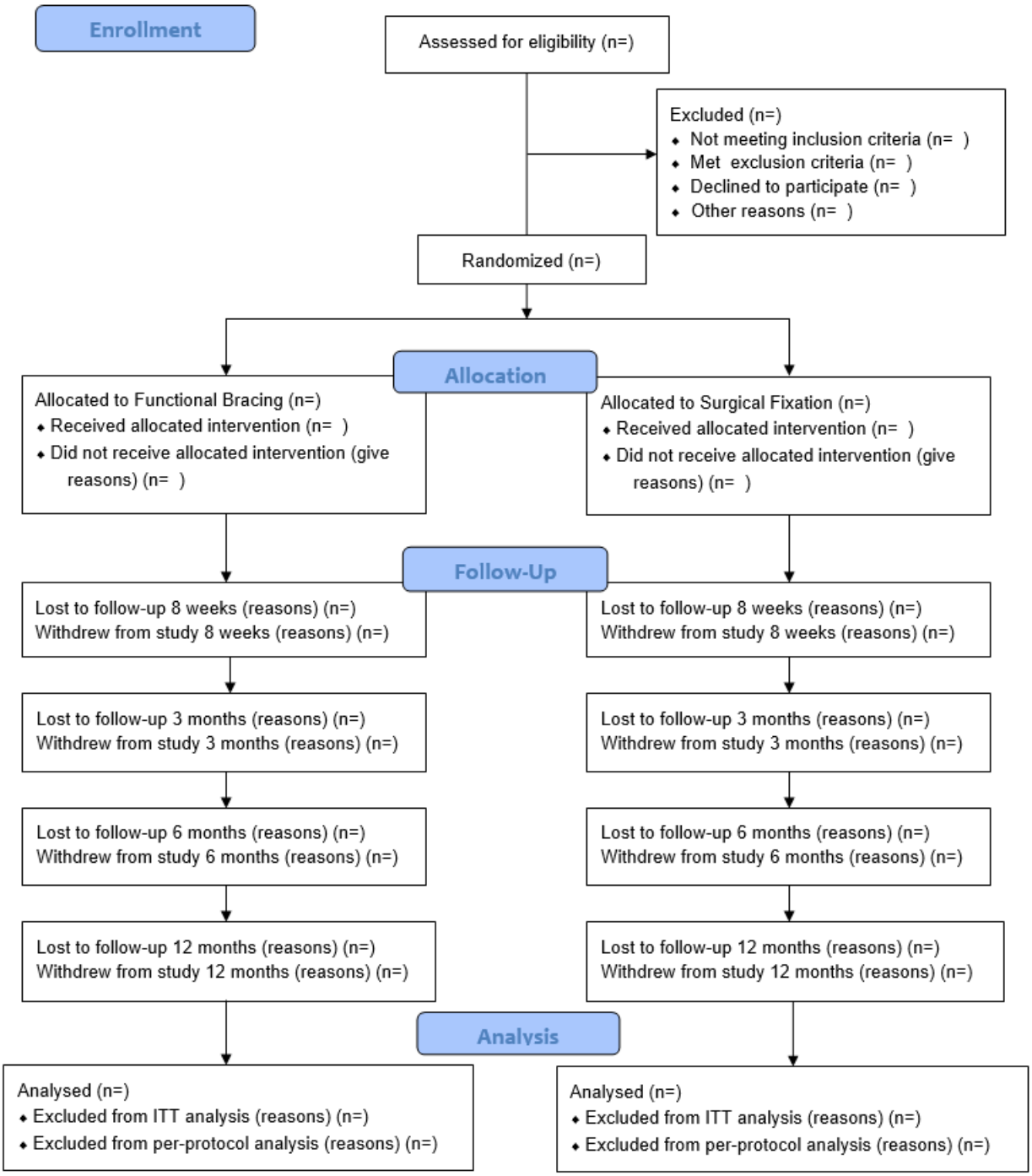

Table 2 Reasons for exclusion

REASON FOR EXCLUSION	DDMonYYYY-DDMonYYYY N (%)
Not meeting inclusion/exclusion criteria	N=
Inclusion criteria 1 Fracture of the humeral diaphysis which the surgeon believes may benefit from surgical fixation	

Inclusion criteria 2 'Diaphysis' defined as the section of bone outside 1 Muller-square of the proximal and distal ends of the Humerus	
Inclusion criteria 3 Participant is willing and able to give informed consent for participation in the study.	
Inclusion criteria 4 Adults, aged 18 years or above.	
Exclusion criteria 1 The fracture is open	
Exclusion criteria 2 The fracture is complicated by local tumour deposits	
Exclusion criteria 3 Bilateral fractures	
Exclusion criteria 4 The index injury occurred more than 16 days prior to recruitment	
Exclusion criteria 5 The patient is unable to adhere to trial procedures	
Exclusion criteria 6 Other upper limb injuries which may reasonably be expected to affect responses to outcome PROMs	
Total	
Eligible but not randomised	N=
Missed	
COVID-19	
Eligible but refused	
Treatment preference	
Non-op	
Surgery	
No research	
No questionnaires	
Other	
No reason given	
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 deviation

	Functional bracing (N =)	Surgical fixation (N =)	Total (N =)
Number of protocol deviation			
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 number (and percentage) of withdrawal reasons will be presented by treatment arm detailed in the **HUSH_SAPDefinitionTables_V2.1_28Oct2024**. Details of each withdrawal will be investigated, including their days on the trial, site, last questionnaire completed and whether they have their primary outcome available for analyses. These will be summarised in Table 4 and Table 5. Any deaths (and their causes) will also be reported in a separate Table as detailed in the **HUSH_SAPDefinitionTables_V2.1_28Oct2024**.

Table 4 Withdrawal details

Participant ID	Time to withdrawal (Days)	Centre	Withdrawal Reason	Last completed questionnaire	Primary outcome available
ID 1					Y/N
ID 2					Y/N

Table 5 Summaries of withdrawals

	Functional bracing (N =)	Surgical fixation (N =)	Total (N =)
Number of withdrawals			
Primary outcome available			
Withdrawal type			
Withdrawal from completing any further questionnaire			
Withdrawal from questionnaires AND			

routine NHS data collection
Withdrawal reason
Participant doesn't like the idea of being part of research
Participant doesn't want to complete questionnaire
No reason given
Other reason

5.3 Baseline Characteristics

Baseline comparability between two treatment groups (functional bracing and surgical fixation) will be investigated. The treatment groups will be compared in terms of stratification factors and baseline characteristics (see **HUSH_SAPDefinitionTables_V2.1_28Oct2024**). Numbers (with percentages) for binary and categorical variables and means (and standard deviations), or medians if the data is skewed (with lower and upper quartiles) 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.

5.4 Unblinding

This is an unblinded study.

5.5 Treatment Compliance with Details of Intervention

A summary of treatment received (functional bracing, surgical fixation or other treatment) will be provided by treatment arm. Non-compliance to the allocated treatment will be summarised by treatment arm together with their reasons (clinical decision, participant decision, lack of equipment, administrative error, or other reason) (see **HUSH_SAPDefinitionTables_V2.1_28Oct2024**). The table will also summarise the number of participants switching to another treatment arm. Participants randomised to functional bracing, and who receive surgery within 6 weeks after randomisation, will be included among those who switched and reported. The reporting of treatment switching for those allocated to functional bracing will distinguish between those who switched immediately and those who switched later. Participants randomised to surgery, who receive functional bracing, are also considered to be switching treatments.

The fraction of exclusions associated with the definition of the per-protocol population will be reported by arm as well as baseline comparability of participants included in the per-protocol analyses.

5.6 Reliability

To ensure consistency, validation checks of the data will be conducted. This will include checking for duplicated records, checking the range of variable values and validating potential outliers and anomalous results where possible (query 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 and data collection instruments have been validated prior to data entry commencing. Calculations and processes performed by a computer program, including the construction of derived data, will be checked These checks will also confirm whether the data has been imported into the statistical software correctly and will check any merging of different datasets, and calculations of derived variables. Clarification will be sought from the trial office in the case of discrepancies.

For each variable, missing values will be checked for consistency and proportion of missing values per variable will be assessed. Patterns of missing data across treatment arm and stratification factors will be explored.

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 and/or identified with major protocol deviations in the blinded review.

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 [9] and described in section 6.6 supplementary analyses addressing a secondary estimand.

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 DASH scores (ATE) between the two interventions in the combined groups randomised to receive each intervention. **The resulting treatment effect represents the average difference in DASH scores we can expect if all participants in the target population were treated with functional bracing instead of being treated with surgical fixation** [10-12].

Supplementary analyses (section 6.6) focus on a secondary estimand under a principal stratum strategy and targeting the treatment effect on those treated with surgical fixation (or complier average treatment effect CACE) and **estimates the contrast in DASH scores between those who received surgical fixations and those treated with functional bracing who could have received surgical fixation had they been randomised to surgical fixation**. The primary estimator for this estimand will be implemented via an instrumental variable approach using randomisation as an instrument. This supplementary analysis will be undertaken If a substantial amount of non-compliance is observed or if the non-compliance is selective.

6.1 Outcome Definitions

6.1.1 Primary outcome

The primary outcome measure for this study is the DASH (Disabilities of Arm, Shoulder and Hand) patient reported outcome measure.

The DASH Outcome Measure is a 30-item, self-reported questionnaire designed to measure physical function and symptoms in patients with musculoskeletal disorders of the upper limb. The items enquire about the degree of difficulty in performing different physical activities because of arm, shoulder and hand problems (21 items), the severity of each of the symptoms of pain, activity-related pain, tingling, weakness and stiffness (five items) and the impact of the problem on social functioning, work, sleep and self-image (four items). Each item has five response options. The scores are then used to calculate a scale score ranging from 0 (no disability) to 100 (most severe disability)—this is called the DASH score [13]. The questionnaire was designed to help describe the disability experienced by people with upper-limb disorders and also to monitor changes in symptoms and function over time. Testing has shown that the DASH performs well in both these roles. DASH has been the most consistently reported Patient Reported Outcome Measure (PROM) in studies investigating humeral fractures and as such, its use will allow contextual comparison with previous and future work.

Previous work has highlighted the reliability and validity of the DASH score in the study of humeral fractures [14].

6.1.2 Secondary outcomes

Pain VAS: To assess pain recovery in the immediate post-injury period (up to week 8), a visual analogue scale (VAS) on a scale of 0-100, where 0 is no pain at all and 100 is the worst pain imaginable, will be used [15]. This will be administered through SMS/text message or email.

DASH Sports/Performing arts module: This additional sub section of the DASH questionnaire investigates the effect of upper limb injury on a patient's participation in sports or playing instruments. The measure consists of 4 questions which will be completed if the participant indicates usual participation in sports or musical activities. The questions are each scored on a 5-point Likert scale. This module ranges from 0 (not disabled) to 100 (most severe disability) [13].

PROMIS Physical Function (upper extremity) and PROMIS Pain Interference: Patient Reported Outcome Measurement Information System (PROMIS) questionnaires are patient reported outcome measures which are administered electronically [16]. They represent a form of Computer Adaptive Test (CAT). CATs are dynamic tests based on Item Response Theory (IRT), a mathematical model that adapts the sequential questions based on a participant's previous response, enabling the successive administration of a tailored set of questions from a large item pool. CATs have been validated in a variety of chronic health conditions. Multiple instruments have been designed including the United States (US) National Institute of Health, PROMIS. PROMIS instruments cover a variety of domains, and are scored from 0 to 100 with 50 points representing the mean score for the US general population and higher scores indicate better function. These instruments address the demand for shorter, more practical measurement of patient-focused outcomes with increased efficiency and precision. This study will utilise the Physical Function (upper extremity) which focusses on function and disability and the pain-interference PROMIS questionnaires which investigates pain intensity and impact. Both of these questionnaires have been found to be valid in the context of upper limb fractures [17, 18]. If internet access is not available to the participant, paper-based (short-form) versions of the PROMIS questionnaires will be sent to the participants for completion.

EQ-5D-5L: The EuroQol 5 Dimensions (EQ-5D-5L) is a validated, generalised and standardised instrument comprising a VAS measuring self-rated health and a health status instrument, consisting of a five-level response for five domains related to daily activities; (i) mobility, (ii) self-care, (iii) usual activities, (iv) pain and discomfort and (v) anxiety and depression [19]. For instance, for mobility the options are: no problems, some problems, moderate problems, severe problems and unable. Responses to the health status classification system are converted into an overall score using a published utility algorithm for the UK population. A respondent's EQ-VAS gives self-rated health on a scale where the endpoints are labelled 'best imaginable health state' (100) and 'worst imaginable health state' (0). We will follow the most up-to-date position statement from NICE when processing the data. Utility scores for the UK population will be used to derive 12 months quality-adjusted life years (QALYs) using the area under the curve method [20].

Resource use: Patient and hospital reported resource use we will be recorded including hospital admissions, outpatient appointments and personal social services use in relation to injury.

Return to work: As the time to return to work is such an important aspect of patient management, this specific element of cost effectiveness will be explored by weekly SMS/email.

Complications: All complications will be recorded, but particular note will be made of complications related to the surgical procedure (wound infection, nerve injury, injury to a blood vessel, non-union, shoulder stiffness, elbow stiffness), and problems identified during the Patient and Public Involvement (PPI) process by those having undergone functional bracing (pressure sores, elbow stiffness). Radiographic images obtained as

part of routine practice will also be collected. In particular, we will seek images pre-operatively, post-treatment, and as close as possible to 12 months (the primary outcome) following randomisation.

Details of all derivations associated with all primary and secondary outcomes can be found in the **HUSH_SAPDefinitionTables_V2.1_28Oct2024.**

6.2 Analysis Methods

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

Unadjusted summary statistics of the DASH 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 [21]. The model will include fixed effects to adjust for the stratification factors used by the randomisation system: recruitment centre, age and nerve injury at presentation. 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 DASH questionnaire completed.

θ denotes the overall treatment 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:

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

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

MVN denotes a multivariate normal distribution quantifying the between-individual variation in means and linear change over time, respectively along with their correlation and $\varepsilon_{i,t} \sim Normal(0, \sigma_{i,t}^2)$ 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 [22] 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 [23, 24].

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 DASH 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 [25]. 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 [26] 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 [27]. 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. [28, 29]. The inverse probability weighting scheme for the propensity of receiving the randomised treatment will be modelled as a function of baseline covariates including subgroup membership, type of surgery instrumentation, age, sex, stratification factors, type of fracture, whether the fracture is simple or complex, BMI, diabetic status and presence of concomitant injuries which affect limb function. All interactions between group membership and the rest of the baseline covariates will be included in this adjustment set [30].

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 [31] conditional on all outcome measurements on all other visits and other covariates included in the model.

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 6). 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.

Table 6: Patterns of Missing data and drop out

Observation Occasion	**Treatment group: Randomised to Functional Bracing**				
	N available	**Missed occasion**		**Dropped-out**	
		N_m	**%**	**N_d**	**%**
Baseline					
8 weeks					
3 months					
6 months					
12 months					

	Treatment group: Randomised to Surgical Fixation				
	N available	Missed occasion		Dropped-out	
		N_m	%	N_d	%
Baseline					
8 weeks					
….					

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 DASH scores as well as follow-up DASH 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 [32] (e.g. using xteregress and/or GLLAMM in Stata) where the longitudinal DASH scores will be modelled jointly with the drop-out indicator. The analysis will be performed alongside for the main ITT estimand only. If treatment switching is found to be associated with drop out then this analysis will also be repeated for the secondary estimand (ATT).

6.4 Sensitivity Analysis

Primary estimand (ATE)

Supplementary analyses will be undertaken for the primary estimand [9] where all randomised patients will be analysed according to their treatment allocation and adjusting for treatment switching using two different approaches for the primary estimand (ATE): **Propensity Score (Quantile) Regression adjustment and Inverse Probability of Treatment weighting (IPTW).** 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 participants has a nonzero probability to be treated with surgical fixation (positivity assumption) [11, 12, 33] .

The initial step in both propensity regression adjustment and IPTW methods 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 [34, 35] but can be sensitive to model misspecification. 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.

In the IPTW approach [36], the PS is used to create inverse probability weights, to construct the classical Horvitsz-Thompson survey estimator (unstabilised weights) whereby among the treated (randomised to surgical fixation) we up-weight those who had a low probability of receiving surgical fixation (participants under-represented in the surgical fixation group) and among the untreated (those randomised to functional bracing) we up-weight those who were unlikely to receive functional bracing. Stabilised weights for the IPTW estimator will be used to increase statistical efficiency and attain better coverage of confidence interval as recommended Stabilised weights should have a mean of 1 and further trimming of the weights will be applied if necessary. The approach is less sensitive to model misspecification but tends to produce estimates with reduced precision [37, 38].

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.

Both of the above estimators target the average difference in DASH scores between the two groups (ATE) using weights constructed from the combined sample of treated with surgical fixation and control participants (treated with functional bracing) as the target population to which each is standardised. **The resulting treatment effect represents the average difference in DASH scores we can expect if all participants in the target population were treated with surgical fixation instead of being treated with functional bracing** [10].

Missing outcome data: The primary analysis method proposed is reasonably robust to missing at random (MAR) data [31]. 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. [39], 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 DASH scores).

6.5 Pre-specified Subgroup Analysis

We will investigate treatment effects across the following subgroups of clinical interest as follows: age, sex, nerve injury at presentation, type of surgery instrumentation (Plate, Nail or brace), fracture position (middle 1/3 fracture, proximal 1/3 fracture or distal 1/3 fracture), whether the fracture is simple or complex, concomitant injuries at presentation which affect limb function and BMI. 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 [6].

If there are too few participants in any subgroup (<=15) or one treatment arm of a subgroup (<=5), the analysis 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 examined in separate models for each clinical grouping of interest. If the clinical grouping is measured in a way that gives rise to a continues 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. The estimated treatment difference and confidence intervals will be reported and presented in forest plots.

6.6 Supplementary/ Additional Analyses and Outcomes

Secondary estimand (ATT): An Analysis under Principal Stratum Strategy:

Participants treated with surgical fixation is an alternative reference population, of interest targeting the treatment effect as a contrast **in DASH scores between those who received surgical fixation and those controls (functional bracing) who could have received surgical fixation had they been randomised to receive surgical fixation.** The primary estimator for this estimand will implement via an instrumental variable approach using randomisation as an instrument (e.g. using xtivreg command in Stata). The instrumental variables approach consists of joint modelling of the longitudinal DASH scores as a function of time and treatment received, jointly with the probability of receiving treatment as a function of the randomised allocation indication (:the instrument) [40] and other variables predictive of receiving treatment. The methods rely on the following assumptions:

a. Ignorability of the instrument. This is likely to hold as the instrument is randomised allocation
b. Non-zero association between instrument and the treatment variable. This is trivially fulfilled and verified empirically from the data.
c. Monotonicity: This assumption tells us that there are no participants who would receive surgical fixation if they were not randomised to surgical fixation but who would not receive functional bracing if they were randomised to functional bracing (we have no “defiers”).
d. Exclusion restriction: Here we assume that the potential outcomes for those participants for whom the receipt of treatment would not be affected by the offer/randomisation to surgical fixation (i.e. would always receive surgical fixation or would receive functional bracing), are no different to the outcomes of those randomised and receiving surgical fixation or functional bracing, respectively. In other words, true compliance is the same as the observed compliance.

The IV approach is an attractive as it does not require the assumption of no unmeasured confounding made in IPTW but the some of the underlying assumptions are arguably tenable, especially the monotonicity restriction maybe violated.

This supplementary analysis will be undertaken If a substantial amount of non-compliance is observed or if the non-compliance is selective.

Informative Missingness: The methods described in Section 6.4 for sensitivity analyses due to informative drop-out will be repeated for the secondary estimand (ATT) if we find that treatment switching is associated with drop-out and outcome scores prior to drop-out. This supplementary analysis will be undertaken only if an ATT analysis is undertaken and there is evidence of informative drop-out.

6.7 Harms

All adverse events are reviewed by the local PI and submitted to the HUSH central office ONLY if they follow 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 7).

Table 7: 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*

Adverse events that are unrelated to the interventions are not reported. Adverse events that are foreseeable and are not SAEs are recorded in the Complications section and are reported as outcomes. Such foreseeable events are described in the protocol section 10.3 (**HUSH_SAPDefinitionTables_V2.1_28Oct2024).**

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 DASH at all time points up to 12 months, and key secondary outcomes: PROMIS Upper Extremity and pain interference at all time points up to 12 months and EQ5D-5L at all time points up to 12 months. Additional variables needed for the validation are stratification factors (recruitment centre, age and nerve injury at presentation), some covariates such as diabetic status, smoking status, body mass index (BMI) and concomitant injuries which affect limb function.

8. SPECIFICATION OF STATISTICAL PACKAGES

All analysis will be carried out using appropriate validated statistical software such as STATA, SAS, SPLUS 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. Franssen, M., et al., *A protocol for the conduct of a multicentre, prospective, randomized superiority trial of surgical versus non-surgical* interventions *for humeral shaft fractures.* Bone & Joint Open, 2024. **5**(4): p. 343-349.
4. Kang, M., et al., *Incorporating estimands into clinical trial statistical analysis plans.* Clinical Trials, 2022. **19**(3): p. 285-291.
5. Kruithof, R.N., et al., *Functional and quality of life outcome after non-operatively managed proximal humeral fractures.* Journal of Orthopaedics and Traumatology, 2017. **18**(4): p. 423-430.
6. 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.
7. Peugh, J.L., et al., *Beyond intent to treat (ITT): A complier average causal effect (CACE) estimation primer.* Journal of School Psychology, 2017. **60**: p. 7-24.
8. Schulz, K.F., D.G. Altman, and D. Moher, *CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials.* BMJ, 2010. **340**: p. c332.
9. 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.
10. 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.
11. 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.
12. 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.
13. 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.
14. Mahabier, K.C., et al., *Reliability, validity, responsiveness, and minimal important change of the Disabilities of the Arm, Shoulder and Hand and Constant-Murley scores in patients with a humeral shaft fracture.* Journal of Shoulder and Elbow Surgery, 2017. **26**(1): p. e1-e12.
15. Hawker, G.A., et al., *Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP).* Arthritis Care & Research, 2011. **63**(S11): p. S240-S252.
16. Rose, M., et al., *The PROMIS Physical Function item bank was calibrated to a standardized metric and shown to improve measurement efficiency.* Journal of Clinical Epidemiology, 2014. **67**(5): p. 516-526.
17. Jayakumar, P., et al., *The use of computer adaptive tests in outcome assessments following upper limb trauma.* The Bone & Joint Journal, 2018. **100-B**(6): p. 693-702.
18. Jayakumar, P., et al., *A Systematic Review of Outcome Measures Assessing Disability Following Upper Extremity Trauma.* JAAOS Global Research & Reviews, 2017. **1**(4).
19. Balestroni, G. and G. Bertolotti, *EuroQol-5D (EQ-5D): an instrument for measuring quality of life.* Monaldi Archives for Chest Disease, 2015. **78**(3).
20. van Hout, B., et al., *Interim Scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L Value Sets.* Value in Health, 2012. **15**(5): p. 708-715.
21. 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.

22. 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.
23. 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.
24. 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.
25. Lesaffre, E., D. Rizopoulos, and R. Tsonaka, *The logistic transform for bounded outcome scores.* Biostatistics, 2006. **8**(1): p. 72-85.
26. Smithson, M. and J. Verkuilen, *A Better Lemon Squeezer? Maximum-Likelihood Regression With Beta-Distributed Dependent Variables.* Psychological Methods, 2006. **11**: p. 54-71.
27. Goldstein, H., *Heteroscedasticity and Complex Variation*, in *Encyclopedia of Statistics in Behavioral Science*. 2005.
28. 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.
29. Zeng, S., et al., *Propensity score weighting for covariate adjustment in randomized clinical trials.* Statistics in Medicine, 2021. **40**(4): p. 842-858.
30. 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.
31. 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.
32. 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.
33. White, I.R., *Uses and limitations of randomization-based efficacy estimators.* Statistical Methods in Medical Research, 2005. **14**(4): p. 327-347.
34. 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.
35. Rubin, D.B. and N. Thomas, *Matching Using Estimated Propensity Scores: Relating Theory to Practice.* Biometrics, 1996. **52**(1): p. 249-264.
36. Hernán, M.A. and S. Hernández-Díaz, *Beyond the intention-to-treat in comparative effectiveness research.* Clinical Trials, 2012. **9**(1): p. 48-55.
37. Basu, D., *An Essay on the Logical Foundations of Survey Sampling, Part One**, in *Selected Works of Debabrata Basu*, A. DasGupta, Editor. 2011, Springer New York: New York, NY. p. 167-206.
38. Horvitz, D.G. and D.J. Thompson, *A Generalization of Sampling Without Replacement From a Finite Universe.* Journal of the American Statistical Association, 1952. **47**(260): p. 663-685.
39. 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.
40. Baltagi, B.H. and Y.-J. Chang, *Incomplete panels: A comparative study of alternative estimators for the unbalanced one-way error component regression model.* Journal of Econometrics, 1994. **62**(2): p. 67-89.

APPENDIX: GLOSSARY OF ABBREVIATIONS

AE	Adverse Event
CI	Chief Investigator
CAT	Computer Adaptive Test
CRF	Case Report Form
CTU	Clinical Trials Unit
DASH	The Disabilities of the Arm, Shoulder and Hand Outcome Measure
DSMC	Data and Safety Monitoring Committee
EQ-5D-Y	EuroQol
HEAP	Health Economic Analysis Plan
GCP	Good Clinical Practice

HE	Health Economy/Economist
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
PROM	Patient Reported Outcome Measure
PROMIS	Patient-Reported Outcomes Measurement Information System
QA	Quality Assurance
QALY	Quality Adjusted Life Year
RCT	Randomised Controlled Trial
REC	Research Ethics Committee
REDCap	Research Electronic Data Capture
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SAR	Serious Adverse Reaction
SD	Standard Deviations
SFQ	Site Feasibility Questionnaire
SOP	Standard Operating Procedure
SAP	Statistical Analysis Plan
SAR	Serious Adverse Reaction
SUSAR	Serious Unexpected Adverse Reaction
TMG	Trial Management Group
TSC	Trial Steering Committee
VAS	Visual Analogue Scale
WPAI	Work Productivity and Activity Impairment Questionnaire