BJO

■ CHILDREN'S ORTHOPAEDICS

The Forearm Fracture Recovery in Children Evaluation (FORCE) trial: statistical and health economic analysis plan for an equivalence randomized controlled trial of treatment for torus fractures of the distal radius in children

R. Knight,
M. Dritsaki,
J. Mason,
D. C. Perry,
S. J. Dutton

University of Oxford, Oxford, UK

Aims

Torus fractures of the distal radius are the most common fractures in children. The NICE non-complex fracture guidelines recently concluded that bandaging was probably the optimal treatment for these injuries. However, across the UK current treatment varies widely due to a lack of evidence underpinning the guidelines. The Forearm Fracture Recovery in Children Evaluation (FORCE) trial evaluates the effect of a soft bandage and immediate discharge compared with rigid immobilization.

Methods

FORCE is a multicentre, parallel group randomized controlled equivalence trial. The primary outcome is the Wong-Baker FACES pain score at three days after randomization and the primary analysis of this outcome will use a multivariate linear regression model to compare the two groups. Secondary outcomes are measured at one and seven days, and three and six-weeks post-randomization and include the Patient Reported Outcome Measurement Information System (PROMIS) upper extremity limb score, EuroQoL EQ-5D-Y, analgesia use, school absence, complications, and healthcare resource use. The planned statistical and health economic analyses for this trial are described here. The FORCE trial protocol has been published separately.

Conclusion

This paper provides details of the planned analyses for this trial, and will reduce the risks of outcome reporting bias and data driven results.

Introduction

Torus (buckle) fractures of the distal radius are the most common fractures in children, accounting for 500,000 UK emergency attendances annually.[1] Children's bones are very flexible compared with adult bones, and can 'buckle' such that there is deformation but no break in the cortex. Torus fractures are very low risk injuries for complication or deformity in the skeletally immature, and these fractures universally heal well.[2]

There is considerable variation in the management of torus fractures, with treatment varying from the use of a removable splint, to plaster cast immobilization, to more flexible splints. The variation in practice has arisen from a longstanding taught doctrine of rigid immobilization for fractures,[3] tempered with newer evidence to suggest that simpler treatment methods are frequently as effective or perhaps even more effective.[4-8] Two systematic reviews support the abandonment of non-removable rigid casts in favour of splints removable at home.[9,10] The first of these identified eight studies, including 825 participants in total, and highlighting several benefits of alternative types of splints over casts.[9] The second review included eight randomized controlled

Correspondence should be sent to Ruth Knight; email: ruth.knight@csm.ox.ac.uk

trials (RCTs) (781 participants in total), demonstrated a variety of benefits of non-rigid immobilization, and showed patients would like to choose non-rigid methods of immobilization for future use.[10]

The recent NICE non-complex fracture guidelines made recommendations on the management of these injuries,[1] concluding that torus fractures of the distal radius should not be immobilized in a non-removable rigid cast, and advocating discharge from the emergency department without subsequent need for outpatient follow-up. NICE recommended that bandaging or soft casts should be the mainstay of treatment for torus fractures but questioned whether any intervention was necessary at all.

The best evidence for minimal immobilization for the treatment of torus fractures of the distal radius comes from the use of soft bandaging in two small, low-quality randomized trials comparing soft bandaging to rigid plaster casts.[11,12] These trials identified improved pain, improved function, less school absence and better convenience with bandaging, although there was a degree of parental anxiety about not using a splint/cast to treat a broken bone. There have been no trials comparing removable rigid splints to bandage treatment.

The FORCE study is a randomized equivalence trial of soft bandage versus immobilization with splints as per current practice. The protocol paper for FORCE is also published separately in this journal;[13] the aim of this paper is to report in detail the statistical analysis plan as reviewed by the trial steering committee and the plan for the health economics analysis. This paper has been prepared according to the published guidelines on the content of statistical analysis plans[14] and the guidance on how to conduct economic evaluations alongside RCTs.[15,16]

Methods

Trial design. FORCE is a multicentre, two-arm, parallel group, randomized, controlled, equivalence trial comparing offer of a soft bandage and immediate discharge with splint immobilization and follow-up as per the protocol of the treating centre for the treatment of torus fractures of the distal radius in children. Children aged four to 15 years who have sustained a torus fracture within the last 36 hours, show no cortical disruption on radiographs, and have no additional fractures (other than ipsilateral ulna fractures) are eligible for this study. Eligible patients are randomized on a 1:1 basis to the two treatment groups. Randomization is stratified by recruitment centre and by age group (four to seven-year-olds versus eight to 15-year-olds). The nature of treatments used means that participants, their parents/guardians, and their treating clinicians cannot be blind to treatment allocation. The primary outcome is the Wong-Baker FACES pain rating scale[17] measured at three days post-randomization. The feasibility of collecting this outcome electronically was established in a feasibility study.[18] Secondary outcomes

are assessed at one, three and seven days, and three and six weeks post-randomization. All data collection is performed electronically, and participants and their parents/ guardians are prompted to complete follow-ups via text and/or email. Full details of the trial design, study population, and study procedures are published separately.[13]

Objectives. The primary objective of this trial is to quantify and draw inferences on observed differences in the Wong-Baker FACES pain rating scale at three days post-randomisation between children allocated to soft bandage and immediate discharge and those allocated to rigid immobilization and standard follow-up. Secondary objectives include assessing differences between the same groups in Wong-Baker FACES, Patient Reported Outcome Measurement Information System (PROMIS) upper extremity limb scores, EQ-5D-Y scores, days of school absence, complications, and resource use up to six weeks post-randomization, and differences in regular analgesia use up to seven days post-randomization.

Outcomes

Primary outcome. The primary outcome measure in this trial is the Wong-Baker FACES pain rating scale.[17] This is a validated, self-reported tool consisting of an ordinal assessment of pain using six facial expressions to illustrate the degree of pain intensity (from 0 "no hurt" to 10 "hurts worst") with each face equating to 2 points on the scale. This is self-reported by all participants. This scale will be measured at baseline, one, three, and seven days, and three and six weeks post-randomization, with three days being the primary endpoint.

Secondary outcomes. The secondary outcome measures are as follows:

- PROMIS upper extremity limb score for children:[19,20] this measures the functional recovery of upper limb function. This will be delivered as a computer adaptive test (CAT), with each participant answering an average of eight questions. A CAT enables the answer from one question to inform the choice of the next, so that each child may answer a distinct set of questions to reach their score. PROMIS scores will be self-reported by participants aged eight to 15 years, and proxy-reported for those aged four to seven years. This is measured at baseline, three, and seven days, and three and six weeks post-randomization.

- Analgesia use: whether or not participants have used analgesia in the last 24 hours, and details of the type of analgesia used, will be recorded at one, three, and seven days post-randomization.

- Euroqol EQ-5D-Y:[21,22] the youth version of the EQ-5D-3L, which consists of five domains related to daily activities, each with three possible answers. This will be self-reported by participants aged eight to 15 years, and proxy-reported for the younger age group (four to seven years). It is measured at baseline,

Table I. Summary of outcomes and assessment schedule.

Assessment	Timing					
	Baseline	1 day	3 days	7 days	3 weeks	6 weeks
Demographic details	✓					
Injury details	✓					
Treatment details	✓					
Wong-Baker FACES	✓	✓	✓	✓	✓	✓
PROMIS upper extremity limb score	✓		✓	✓	✓	✓
EQ-5D-Y	✓		✓	✓	✓	✓
Satisfaction		✓				✓
Use of analgesia		✓	✓	✓		
Use of immobilization		✓	✓	✓	✓	
Complications		✓	✓	✓	✓	✓
Days school absence					✓	✓
Resource use					✓	✓

three, and seven days, and three and six weeks post-randomization.

- Days of absence from school/childcare because of the injury: this is measured where applicable. This is proxy-reported for all participants. Data on absence in the preceding 21 days is collected at three and six weeks post-randomization.
- Complications: all complications are recorded, with particular note made of hospital re-attendance for any reason including inadequate analgesia, refracture, or worsening of the fracture. These will be proxy-reported for all participants with additional information provided by clinical teams as necessary.
- Healthcare utilization: NHS costs and patients' parent's out-of-pocket expenses are recorded via a short questionnaire administered at three and six weeks post-randomization.

The data collected at each time point are summarized in Table I.

Sample size

The primary outcome in this study is the six-point Wong-Baker FACES pain rating scale at three days. The Wong-Baker scale has a minimum clinically important difference (MCID) of 1 face which equates to 2 points on the numerical scale.[23] This trial aims to demonstrate the equivalence of a soft bandage compared with rigid immobilization. Assuming an equivalence margin of 1 point (half the MCID), 90% power, one-sided 2.5% significance, and that the standard deviation (SD) is 2.3 (based on a feasibility study), 278 participants (139 per arm) would be required to show equivalence. The Wong-Baker scale is a categorical outcome that is more likely to be non-linear

in younger age groups, tending to linearity in those over eight years of age. [24] Therefore, the trial is powered separately for equivalence in the two subpopulations (four to seven-year-olds and eight to 15-year-olds). In all, 278 participants in the four to seven years age group and 278 participants in the eight to 15 years age group are needed to show equivalence.

Allowing for 20% loss to follow-up inflates the required sample size to 348 in each of the sub-populations (174 per arm). Given the primary outcome is at three days post randomization, the loss to follow-up inflation can be readily adjusted to ensure that the study recruits effectively and efficiently.

Statistical analysis

General analysis principles. Two analysis populations will be considered, the intent-to-treat (ITT) population and the per-protocol (PP) population. The ITT population will include all participants randomized, and participants will be analyzed according to the group to which they were randomized regardless of the treatment they actually received. The PP population will be analyzed according to the treatment they actually received. Participants will be excluded from the PP population if: i) they did not receive the treatment to which they were randomized, or changed from their allocated treatment prior to the primary outcome time point (three days); ii) they did not provide sufficient follow-up data for analysis; or, iii) they did not satisfy the eligibility criteria for the study. Exact exclusion criteria for the PP analysis will be confirmed based on a blinded analysis of the data (not separated by treatment group) prior to the final data lock.

As this is an equivalence trial, a maximum clinical difference (Δ_T) has been prespecified for the primary outcome. This specifies the level within which the two treatments can be considered not to differ in any clinically meaningful way. The null hypothesis we are testing in this trial is that a difference greater than Δ_T exists between the treatments in either direction. The trial is designed to disprove this in favour of the alternative that no clinical difference exists. Analyses of the primary outcome will be carried out on the ITT population and repeated for the PP population and only if both results show equivalence will equivalence be claimed.[25,26]

Analyses of all secondary outcomes will be performed for the ITT population and repeated for the PP population.

The significance level used will be 0.05 with 95% confidence intervals (CIs) reported. All secondary analyses will be considered as supporting the primary outcome analysis and will also be analyzed using a significance level of 0.05 with 95% CIs reported.

To ensure consistency, validation checks of the data will be conducted. As the data is collected electronically, many of these checks will be implemented automatically as part of the data entry procedure. Analysis
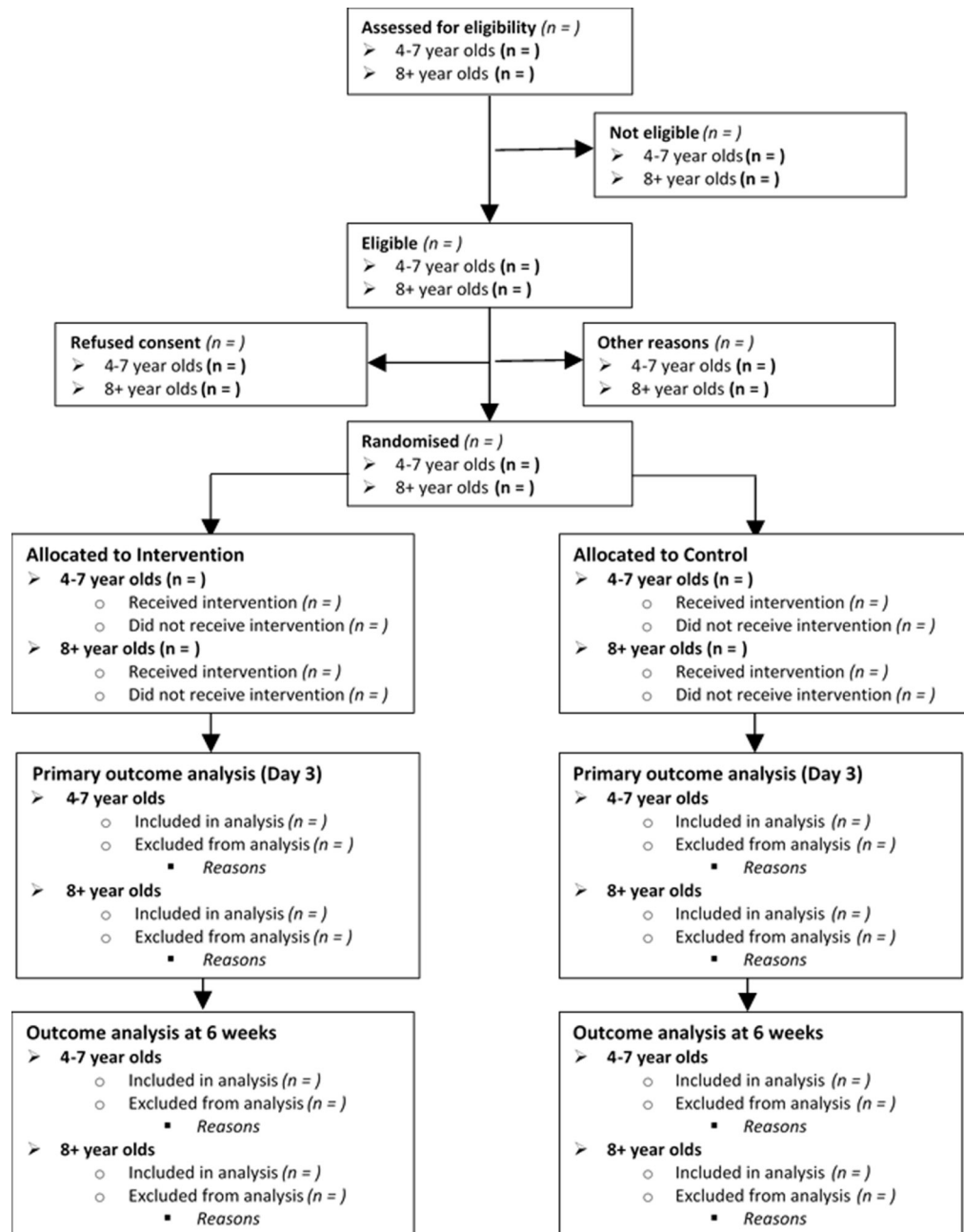
Assessed for eligibility *(n =)*
➢ 4-7 year olds **(n =)**
➢ 8+ year olds **(n =)**

Not eligible *(n =)*
➢ 4-7 year olds **(n =)**
➢ 8+ year olds **(n =)**

Eligible *(n =)*
➢ 4-7 year olds **(n =)**
➢ 8+ year olds **(n =)**

Refused consent *(n =)*
➢ 4-7 year olds **(n =)**
➢ 8+ year olds **(n =)**

Other reasons *(n =)*
➢ 4-7 year olds **(n =)**
➢ 8+ year olds **(n =)**

Randomised *(n =)*
➢ 4-7 year olds **(n =)**
➢ 8+ year olds **(n =)**

Allocated to Intervention
➢ 4-7 year olds **(n =)**
  o Received intervention *(n =)*
  o Did not receive intervention *(n =)*
➢ 8+ year olds **(n =)**
  o Received intervention *(n =)*
  o Did not receive intervention *(n =)*

Allocated to Control
➢ 4-7 year olds **(n =)**
  o Received intervention *(n =)*
  o Did not receive intervention *(n =)*
➢ 8+ year olds **(n =)**
  o Received intervention *(n =)*
  o Did not receive intervention *(n =)*

Primary outcome analysis (Day 3)
➢ 4-7 year olds
  o Included in analysis *(n =)*
  o Excluded from analysis *(n =)*
    ▪ *Reasons*
➢ 8+ year olds
  o Included in analysis *(n =)*
  o Excluded from analysis *(n =)*
    ▪ *Reasons*

Primary outcome analysis (Day 3)
➢ 4-7 year olds
  o Included in analysis *(n =)*
  o Excluded from analysis *(n =)*
    ▪ *Reasons*
➢ 8+ year olds
  o Included in analysis *(n =)*
  o Excluded from analysis *(n =)*
    ▪ *Reasons*

Outcome analysis at 6 weeks
➢ 4-7 year olds
  o Included in analysis *(n =)*
  o Excluded from analysis *(n =)*
    ▪ *Reasons*
➢ 8+ year olds
  o Included in analysis *(n =)*
  o Excluded from analysis *(n =)*
    ▪ *Reasons*

Outcome analysis at 6 weeks
➢ 4-7 year olds
  o Included in analysis *(n =)*
  o Excluded from analysis *(n =)*
    ▪ *Reasons*
➢ 8+ year olds
  o Included in analysis *(n =)*
  o Excluded from analysis *(n =)*
    ▪ *Reasons*

Fig. 1

CONSORT flow diagram for participants in trial up to six weeks follow-up.

of the primary (Wong-Baker FACES) and key secondary (PROMIS) outcomes will be independently repeated by a statistician not involved in the trial to validate the results.

All analyses will be carried out using appropriate, validated statistical software such as STATA[27] or R.[28] The relevant package and version number used for the analysis will be recorded and reported.

Descriptive analyses. The flow of participants through each stage of the trial, including the number of individuals screened, eligible, randomized to each arm, receiving allocated treatment, and included in the primary analysis will be summarized using a CONSORT flowchart (Fig. 1). Reasons for ineligibility, loss to follow-up and exclusion from the primary analysis will be summarized. The data will be summarized separately for the two age groups (four- to seven-year-olds and eight- to 15-year-olds).

The baseline comparability of the two randomized groups in terms of: i) stratification factors; ii) baseline characteristics (Table II); and iii) primary and secondary outcomes at baseline, will be presented. Numbers with

Table II. Baseline characteristics.

Baseline characteristic	Type	Levels or scale
Sex	Binary	Male; Female
Age	Continuous	Years
Side of injury	Binary	Right; Left
Mechanism of injury	Categorical	Low energy fall; High energy fall; Other
Dominant hand	Categorical	Right; Left; Unsure/ambidextrous

percentages will be used to compare binary and categorical variables, and either means and SDs or medians and interquartile ranges (IQRs) will be used for continuous variables. There will be no tests of statistical significance nor CIs for differences at baseline between the randomized groups.

Loss to follow-up, withdrawals, and missing data. The numbers (and percentages) of losses to follow-up and withdrawals along with reasons for these will be reported by treatment group at each time point. If differential losses are identified the reasons for these will be explored further. No deaths are anticipated in this study; however, if any do occur these will be reported separately together with reasons.

The number and percentage of individuals with missing data for each outcome at each time point will be summarized by treatment group, along with reasons for missing-ness if known. The patterns of missing-ness will also be explored and the suitability of missing data assumptions considered.

All the main analyses in this trial will be performed using the available case dataset. Since the primary outcome time point is three days after randomization, it is not anticipated that there will be a high volume of missing data. In addition, the multi-level model using repeated measures which is planned as one of the supporting analyses of the primary outcome is relatively robust to missing data under the missing at random (MAR) assumption.[29] However, if more than 10% of the data is missing, then sensitivity analyses exploring a variety of missing not at random scenarios will be undertaken to ensure the results are robust to departures from the MAR assumption. No imputation is planned for any of the secondary outcomes.

Compliance. The numbers and percentages of participants who: i) received their allocated treatment (splint immobilization or offer of a bandage) throughout/until removal of treatment; ii) received another treatment at baseline; iii) received their allocated treatment at baseline but changed to another treatment prior to the primary outcome time point (three days); and, iv) received their allocated treatment at baseline but changed to another treatment after the primary outcome time point, will be summarized by treatment group. Reasons for not receiving the allocated treatment, or for changing from the allocated treatment will also be summarized.

For those receiving splint immobilization (either as their allocated treatment or as a deviation from their allocated treatment), the number and percentage receiving each type of immobilization will be summarized as well as the average length of time the splint is worn. For those who are offered a bandage, details on the number and percentage who have this applied prior to discharge and who take it home and apply it at home will be summarized. The average length of time the bandage is worn will also be summarized.

Analysis of primary outcome. Wong-Baker FACES scores at three days after randomization will be summarized by treatment group using means and SDs. A multivariate linear regression model adjusting for stratification factors (age and recruitment centre) and participant sex will be used to compare the two groups. The appropriateness of the assumption of approximate normality of the residuals of this model will be assessed graphically. The adjusted difference between treatment groups will be reported along with associated 95% CIs and p-values. This analysis will be performed for the ITT population and repeated for the PP population both using available cases, and only if both results show equivalence will equivalence be claimed. An unadjusted *t*-test will also be performed.

These analyses will be repeated separately for the two age groups (four to seven-year-olds and eight to 15-year-olds) with results reported in a similar manner. Separate analyses for the two subgroups, instead of the inclusion of an interaction effect, will allow the treatment effects in the two subgroups to be considered separately. This may be important because the Wong-Baker score may behave differently in the two age groups; non-linear in younger children but tending to linearity in those over eight. The trial has been powered accordingly.

In addition, the analysis will be repeated to explore the impact of receipt of pain medication prior to this time point. This will involve repeating the analysis separately for the subgroups, or using seemingly unrelated regression[30] as appropriate.

Analyses utilizing all time points (baseline to six weeks) will also be undertaken. Multi-level linear regression models using repeated measures (level 1) nested within participants (level 2) and adjusted for recruitment centre (level 3) and participant sex and age (fixed effects) will be used. Trends over time will be examined, and, if appropriate, interactions between treatment and time will be included in the model. The appropriateness of the assumption of approximate normality of the residuals of the model will be assessed graphically. Wong-Baker FACES scores at each time point will be summarized by treatment group using means and SDs and the adjusted difference between these will be reported along with the associated 95% CI.

In addition, supplementary analyses using the area under the curve (AUC) summary statistics[31] will be

conducted to investigate total pain: i) from baseline to three days after randomization; and ii) from baseline to six weeks after randomization. Parameter estimates from the mixed effects model will be used to calculate AUCs for each treatment group providing an overall estimate of pain over each time period (up to three days and up to six weeks). These will be presented with associated 95% CIs. The difference between the groups will be calculated and compared using a *t*-test.

If the assumption of approximate normality is not appropriate, the first approach will be to consider a transformation of the data to attain normality. If normality cannot be achieved by transformation, the data will be analyzed separately for each time point using a non-parametric equivalent with no adjustment and medians and IQRs will be reported for each treatment group.

Analysis of secondary outcomes. Continuous secondary outcomes (PROMIS and EQ-5D-Y utility and VAS) will be analyzed using repeated measures mixed effects multilevel linear regression models similar to that outlined above. Scores at each time point will be summarized by treatment group using means and SDs and the adjusted difference between these will be reported along with the associated 95% CI. If the assumption of approximate normality is found not to be appropriate, the methods outlined earlier will be implemented.

The number and percentage of participants using pain medication in the first 7 days after randomization will be summarized by treatment group. A mixed effects logistic regression model adjusted for recruiting centre and participant sex and age will be used to compare the two treatment groups, and the adjusted odds ratio (OR) and associated 95% CI will be reported. Risk differences and associated 95% CIs will also be reported. Trends over time by treatment group will be presented graphically. Any participants receiving something other than standard 'over-the-counter' pain-killers will be summarized.

The number and percentage of participants who have missed school up to 6 weeks after randomization will be similarly summarized and analyzed. The number of days of school missed will also be summarized by treatment group using medians and IQRs, and the two arms will be compared using non-parametric methods.

Since PROMIS scores, EQ-5D-Y scores and use of pain medication are proxy-reported for the younger age group and patient-reported for the older age group, the analyses of these outcomes will be repeated including an interaction between treatment group and age group. Models analogous to those outlined previously will be used.

Analysis of complications and SAEs. The number and percentage of participants experiencing a foreseeable complication (recall to hospital with a diagnosis of an alternative fracture pattern or worsening fracture deformity) will be summarized by treatment group. The total number of complications experienced in each arm will also be summarized. If there are sufficient numbers of complications, a mixed effects logistic regression model adjusted for recruiting centre and participant age and sex will be used to compare the rates of complications between treatment groups and the odds ratio and 95% CI will be reported. Risk differences and associated 95% CIs will also be reported. The number of participants experiencing each type of complication will also be summarized by treatment group and, if there are sufficient numbers of any type of complications, the two groups will be compared using a chi-squared test. If sufficient numbers of complications occur, temporal patterns will be explored graphically.

The number and percentage of participants experiencing an SAE in each treatment group will be summarized and compared in a similar manner. The total number of SAEs in each arm will also be summarized.

Supplementary analysis. Within the rigid immobilization group there are two types of immobilization which can be used, splint or cast. An exploratory analysis investigating the effect of type of immobilization used on Wong-Baker scores at three days after randomization will be conducted. This analysis will only be performed if a sufficient number of participants receive each type of immobilization (at least 10% to each type).

Health economics analysis

A within-trial cost-utility analysis with imputation will be conducted from an NHS and Personal Social Services (PSS) perspective using the FORCE study data over the period of six weeks.

Measurement of resource use. Resource use data for the economic evaluation will be collected during the trial period from online questionnaires sent to participants at three and six weeks after randomization. Questionnaires will capture resource use associated with torus fractures of the distal radius: included are the frequency of use of outpatient care, community care and social care services. Questionnaires will also record private care (e.g. physiotherapy), direct medical costs (e.g. medications), direct non-medical costs (e.g. help with childcare) and indirect costs (i.e. carer absenteeism) associated with the injury. These health resource questionnaires will be completed by the participant, or on behalf of the participant (the proxy, usually the participant's parent/guardian). The questionnaires will cover two survey periods; baseline to three weeks, and three weeks to six weeks post-randomization.

Valuation of health outcomes. An EQ-5D-Y value set is currently not available for the UK.[32] We will value EQ-5D-Y responses using the most appropriate valuation set available for the trial population at the time of analysis. A recent review of patterns and trends of measurement and valuation of childhood health utilities[33] reveals that 78.7% of the studies using EQ-5D-Y used general

adult-derived tariffs whereas the rest did not provide any information about the tariff used. If necessary the adult EQ-5D-3L tariff will be applied, in which case we will undertake sensitivity analysis to make sure that trial findings are not sensitive to the valuation set chosen.[34] Quality-adjusted life years (QALYs) will be calculated as the area under the utility curve of EQ-5D utility scores recorded at baseline, three, and seven days, and three and six weeks post-randomization data using the trapezoidal rule.[35]

Valuation of direct medical costs. Unit costs for resource items will be sourced from the latest national sources such as the NHS Supply Chain catalogue[36] and NHS reference costs.[37] Unit costs for the different forms of immobilization applied (futura splint, back-slab, soft cast full, soft cast splint, hard cast full and hard cast splint) and gauze bandage will be sourced from the latest NHS Supply Chain catalogue. Since the injury is primarily managed solely within the Accident and Emergency department, the cost of hospitalisation as defined by the Health Resource Group (HRG) code is expected to be the same between the trial arms, as will be the unit cost for labour, the number and type of staff and duration of treatment. These data will be reported as part of the FORCE trial and included in the cost-utility analysis.

Sequelae resource use following trial interventions, such as outpatient care and community care, will be costed using the latest available NHS reference costs[37] and Personal Social Services Research Unit (PSSRU) costs.[38] The unit cost of medication related to wrist injury will be sourced using the latest available British National Formulary (BNF) and the Prescription Cost Analysis (PCA) for England.[39] Costs of medications for individual participants will be estimated based on their reported doses and frequencies, when these are available, or reasonable average values when partially reported.

Valuation of direct non-medical and indirect costs. Patient non-medical costs will be obtained directly from the patient questionnaire. Parent or carer time off from work will be estimated using the daily median wage reported by the Office for National Statistics.[40] These costs will be excluded from the base case analysis, limited to the NHS/PSS perspective of the economic evaluation.

The patient-level cost of health resource use will be calculated by multiplying resource utilization by its unit cost. The base currency of all costs will be the most recent year for which unit cost data are available and expressed in UK pounds (£).

Analysis of health economic outcomes. Within most trials some incompleteness of data is anticipated. Assuming missingness is non-trivial (5% or more), the base case analysis will use multiple imputation using the MI suite of commands in Stata.[27] The imputed within-trial incremental cost and quality-adjusted life years (QALYs) gained will be the base case, adjusted for trial stratification covariates.

Complete case analysis will be reported as a sensitivity analysis.

Imputation will follow good practice guidance[41,42] exploring the extent to which data are missing at random, for example by using logistic regression for missingness of costs and QALYs against baseline variables.[43] Multiple imputed datasets (or 'draws') will be generated by treatment group, with missing values informed by regular (trial covariates) and imputed variables (costs and EQ-5D-Y at each time point). To improve the robustness of imputed values, predictive mean matching drawn from the five nearest neighbours will be used. The imputation model will use fully conditional (MCMC) methods (multiple imputation by chained equations), which manage missing and correlated data in multiple dependent variables. Each draw will be analyzed independently using bivariate regression within the Stata MI suite (StataCorp, College Station, Texas, USA), providing estimates adjusted for within and between variances for imputed samples.[44] The number of draws will be set to minimize the information loss of finite imputation sampling – typically 20 draws. Imputed and observed distributions will be compared visually and statistically to establish the consequences of estimation.

Within the MI framework, seemingly unrelated regression equations (SUREG) will be used to model bivariate incremental changes in costs and QALYs. Baseline QOL scores will be included within models to allow for potential baseline imbalances.[45] Joint distributions of costs and outcomes will be generated using the (non-parametric) bootstrap method, with replicates used to populate a cost-effectiveness plane. Using bias-corrected non-parametric bootstrapping, 5000 bootstraps will be taken per model or draw evaluated. Mean estimates will be reported with 95% CIs.

The incremental cost-effectiveness ratio (ICER) will be estimated as the difference between treatments in mean total costs divided by the difference in mean total QALYs within six weeks from intervention. Value-for-money is determined by comparing the ICER with a threshold value, typically the NICE threshold for British studies, of £20,000 to £30,000/QALY.[46] Sensitivity analyses will include an exploration of treatment interaction with the trial age strata.

The net monetary benefit (NMB) of changing treatment will be reported as a recalculation of the ICER at a range of thresholds of willingness to pay for an additional QALY. The NMB describes the resource gain (or loss) when investing in a new treatment at a given willingness to pay threshold. NMB estimates will be used to generate cost-effectiveness acceptability curves (CEACs). The CEAC compares the likelihood that treatments are cost-effective as the willingness to pay threshold varies.[47]

The expected value of perfect information (EVPI) is the upper limit of the value to a healthcare system of further research to eliminate uncertainty.[48] Findings from

cost-effectiveness analyses remain uncertain because of the imperfect information they use. The NMB framework allows this expected cost of uncertainty to be determined and guide whether further research should be conducted to eliminate uncertainty.

Further cost analysis will be undertaken from a broader societal perspective. Reporting will follow the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement.[49]

Discussion

The FORCE trial will provide data regarding the effects of offer of a soft bandage on outcomes of participants up to six weeks after sustaining a buckle fracture, compared to those receiving rigid immobilization. This paper provides details of the planned analyses for this trial and will help reduce the risks of outcome reporting bias and data driven results.[50]

Trial status

The trial opened to recruitment on 14 January 2019. Recruitment is currently ongoing, and is expected to finish by June 2020, with follow-up being completed by August 2020. The analysis of all primary and secondary outcomes will be conducted thereafter. These statistical and health economic analysis plans were drafted, reviewed and finalised during the set-up and patient recruitment phases of the trial.

REFERENCES

1. **National Clinical Guideline Centre UK**. Fractures (Non-Complex): assessment and management, 2016.
2. **Solan M**, **Rees R**, **Daly K**. Current management of torus fractures of the distal radius. *Injury*. 2002;33(6):503–505.
3. **Charnley J**. The closed treatment of common fractures: Cambridge university press, 2003.
4. **Davidson J**, **Brown D**, **Barnes S**, **Bruce CE**. Simple treatment for torus fractures of the distal radius. *The Journal of bone and joint surgery British volume*. 2001;83(8):1173–1175.
5. **Symons S**, **Rowsell M**, **Bhowal B**, **Dias JJ**. Hospital versus home management of children with buckle fractures of the distal radius: a prospective, randomised trial. *The Journal of bone and joint surgery British volume*. 2001;83(4):556–560.
6. **Williams KG**, **Smith G**, **Luhmann SJ**, **et al**. A randomized controlled trial of cast versus splint for distal radial buckle fracture: an evaluation of satisfaction, convenience, and preference. *Pediatr Emerg Care*. 2013;29(5):555–559.
7. **Plint AC**, **Perry JJ**, **Correll R**, **Gaboury I**, **Lawton L**. A randomized, controlled trial of removable splinting versus casting for wrist buckle fracture in children. *Pediatrics*. 2006;117(3):691–697.
8. **Oakley EA**, **Ooi KS**, **Barnett PLJ**. A randomized controlled trial of 2 methods of immobilizing torus fractures of the distal forearm. *Pediatr Emerg Care*. 2008;24(2):65–70.
9. **Hill CE**, **Masters JP**, **Perry DC**. A systematic review of alternative splinting versus complete plaster casts for the management of childhood buckle fractures of the wrist. *Journal of Pediatric Orthopaedics B*. 2016;25(2):183–190.
10. **Jiang N**, **Cao Z-H**, **Ma Y-F**, **Lin Z**, **Yu B**, **Z-h C**, **Ma Y-f LZ**. Management of Pediatric Forearm Torus Fractures: A Systematic review and meta-analysis. *Pediatr Emerg Care*. 2016;32(11):773–778.
11. **West S**, **Andrews J**, **Bebbington A**, **Ennis O**, **Alderman P**. Buckle fractures of the distal radius are safely treated in a soft bandage: a randomized prospective trial of bandage versus plaster cast. *J Pediatr Orthop*. 2005;25(3):322–325.
12. **Thapa PB**, **Marasini RP**, **Thapa SK**, **Pradhan NS**, **Giri S**. Soft bandage versus casting in the treatment of Torus fracture of distal radius in children; a prospective randomized study. *Journal of College of Medical Sciences-Nepal*. 2016;11(4):3–8.
13. **Achten J**, **Knight R**, **Dutton SJ**, **Costa ML**, **Mason J**, **Dritsaki M**, **Appelbe D**, **Messahel S**, **Roland D**, **Widnall J**, **Perry DC**. A multicentre prospective randomized equivalence trial of a soft bandage and immediate discharge versus current treatment with rigid immobilization for torus fractures of the distal radius in children: protocol for the forearm fracture recovery in children evaluation (FORCE) trial. *Bone & Joint Open*. 2020;1(6):In press).
14. **Gamble C**, **Krishan A**, **Stocken D**, **et al**. Guidelines for the content of statistical analysis plans in clinical trials. *JAMA*. 2017;318(23):2337–2343.
15. **Petrou S**, **Gray A**. Economic evaluation alongside randomised controlled trials: design, conduct, analysis, and reporting. *BMJ*. 2011;342:d1548.
16. **Ridyard CH**, **Hughes DA**. Methods for the collection of resource use data within clinical trials: a systematic review of studies funded by the UK health technology assessment program. *Value Health*. 2010;13(8):867–872.
17. **Wong DL**, **Baker CM**. Pain in children: comparison of assessment scales. *Pediatr Nurs*. 1988;14(1):9–17.
18. **Widnall J**, **Capstick T**, **Wijesekera M**, **Messahel S**, **Perry DC**. Pain scores in Torus fractures: using text messages as an outcome collection tool. *Bone & Joint Open*. 2020;1(2):3–7.
19. **National Institutes of Health (US)**. PROMIS, Patient-Reported Outcome Meausrement Information System, 2010.
20. **Cella D**, **Yount S**, **Rothrock N**, **et al**. The patient-reported outcomes measurement information system (PROMIS): progress of an NIH roadmap cooperative group during its first two years. *Med Care*. 2007;45(5 Suppl 1):S3.
21. **Eidt-Koch D**, **Mittendorf T**, **Greiner W**. Cross-Sectional validity of the EQ-5D-Y as a generic health outcome instrument in children and adolescents with cystic fibrosis in Germany. *BMC Pediatr*. 2009;9(1):55.
22. **Wille N**, **Badia X**, **Bonsel G**, **et al**. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. *Qual Life Res*. 2010;19(6):875–886.
23. **Garra G**, **Singer AJ**, **Taira BR**, **et al**. Validation of the Wong-Baker FACES pain rating scale in pediatric emergency department patients. *Acad Emerg Med*. 2010;17(1):50–54.
24. **Oliveira A**, **Batalha L**, **Fernandes A**, **Gonçalves J**, **Viegas R**. A functional analysis of the Wong-Baker faces pain rating scale: linearity, discriminability and amplitude. *Revista de Enfermagem Referência*. 2014;IV Série(3):121–130.
25. **Chow S-C**, **Wang H**. On sample size calculation in bioequivalence trials. *J Pharmacokinet Pharmacodyn*. 2001;28(2):155–169.
26. **Christensen E**. Methodology of superiority vs. equivalence trials and non-inferiority trials. *J Hepatol*. 2007;46(5):947–954.
27. **StataCorp L**. Stata Statistical Software: Release 15 College Station, TX. 2017. www stata com/features/documentation/ (date last accessed 1 March 2018).
28. **Team RC**. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria URL. 2017. https://www R-project org
29. **Sullivan TR**, **White IR**, **Salter AB**, **Ryan P**, **Lee KJ**. Should multiple imputation be the method of choice for handling missing data in randomized trials? *Stat Methods Med Res*. 2018;27(9):2610–2626.
30. **Rochon J**. Accounting for covariates observed post randomization for discrete and continuous repeated measures data. *Journal of the Royal Statistical Society: Series B*. 1996;58(1):205–219.
31. **Bell ML**, **King MT**, **Fairclough DL**. Bias in area under the curve for longitudinal clinical trials with missing patient reported outcome data: summary measures versus summary statistics. *SAGE Open*. 2014;4(2):2158244014534858.
32. **Kreimeier S**, **Greiner W**. EQ-5D-Y as a health-related quality of life instrument for children and adolescents: the instrument's characteristics, development, current use, and challenges of developing its value set. *Value Health*. 2019;22(1):31–37.
33. **Kwon J**, **Kim SW**, **Ungar WJ**, **et al**. Patterns, trends and methodological associations in the measurement and valuation of childhood health utilities. *Qual Life Res*. 2019;28(7):1705–1724.
34. **Kreimeier S**, **Oppe M**, **Ramos-Goñi JM**, **et al**. Valuation of EuroQol five-dimensional questionnaire, youth version (EQ-5D-Y) and EuroQol five-dimensional questionnaire, three-level version (EQ-5D-3L) health states: the impact of wording and perspective. *Value Health*. 2018;21(11):1291–1298.
35. **Drummond MF**, **Sculpher MJ**, **Claxton K**, **Stoddart GL**, **Torrance GW**. Methods for the economic evaluation of health care programmes: Oxford university press, 2015.
36. **NHS Business Services Authority**. Nhs supply chain Catalogue 2018/19. *London*. 2018.
37. **Department of Health**. NHS Reference costs 2017 to 2018 [Internet]. https://improvement.nhs.uk/resources/reference-costs/

38. **Curtis L**, **Burns A**. Unit costs of health and social care: PSSRU, Univesity of Kent at Canterbury. *In*. 2018.

39. **Committee JF**, **Formulary BN**. online) [Internet]. *London BMJ Gr. Pharm. Press.* 2019. http://www.medicinescomplete.com

40. **Office for National Statistics**. Labour market overview, UK statistical bulletins [Internet]. 2019. https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/bulletins/uklabourmarket/previousReleases

41. **Sterne JAC**, **White IR**, **Carlin JB**, **et al**. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ.* 2009;338:b2393.

42. **White IR**, **Horton NJ**, **Carpenter J**, **Pocock SJ**. Strategy for intention to treat analysis in randomised trials with missing outcome data. *BMJ.* 2011;342:d40.

43. **Faria R**, **Gomes M**, **Epstein D**, **White IR**. A guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials. *Pharmacoeconomics.* 2014;32(12):1157–1170.

44. **White IR**, **Royston P**, **Wood AM**. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med.* 2011;30(4):377–399.

45. **Manca A**, **Hawkins N**, **Sculpher MJ**. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. *Health Econ.* 2005;14(5):487–496.

46. **National Institute for Health and Care Excellence**. Guide to the Methods of Technology Appraisal. London: NICE, PMG9. 2013. https://www.nice.org.uk/process/pmg9/chapter/foreword

47. **Hoch JS**, **Briggs AH**, **Willan AR**. Something old, something new, something borrowed, something blue: a framework for the marriage of health econometrics and cost-effectiveness analysis. *Health Econ.* 2002;11(5):415–430.

48. **Claxton K**, **Sculpher M**, **Drummond M**. A rational framework for decision making by the National Institute for clinical Excellence (NICE). *The Lancet.* 2002;360(9334):711–715.

49. **Husereau D**, **Drummond M**, **Petrou S**, **et al**. Consolidated health economic evaluation reporting standards (cheers) statement. *BMJ.* 2013;346:f1049.

50. **Finfer S**, **Bellomo R**. Why publish statistical analysis plans? *Crit Care Resusc.* 2009;11(1):5–6.

Author information:
- R. Knight, PhD, Medical Statistician
- M. Dritsaki, PhD, Senior Health Economist
- S. J. Dutton, MSc, University Research Lecturer
  Oxford Clinical Trials Research Unit, Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Botnar Research Centre, Oxford, UK.

- J. Mason, DPhil, Professor of Health Economics, The Centre for Health Economics at Warwick (CHEW), Warwick Clinical Trials Unit, Warwick University, Coventry, UK.
- D. C. Perry, MA (Oxon), PhD, FHEA, FRCS (Tr & Orth), Associate Professor, Oxford Trauma, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Kadoorie Research Centre, Oxford, UK.