eSupplement

Statistical analysis plan

Surgery versus Functional Bracing for Closed Displaced Humeral Shaft Fractures – A 5-year Follow-Up of the FISH Randomized Clinical Trial

Lasse Rämö, Thomas Ibounig, Bakir O. Sumrein, Vesa Lepola, Mika Paavola, Simo Taimela, Teppo L.N. Järvinen, on behalf of Finnish Shaft of the Humerus (FISH) Investigators

This supplement contains the following items:

1. TRIAL REGISTRATION

1.1. Original registration

Original trial registration submitted to Clinical Trials.gov on October 30, 2012 can be accessed at:
https://clinicaltrials.gov/ct2/history/NCT01719887?V_1=View#StudyPageTop

This registration with relevant information is copied below:

Study Start: October 2012 (first patient recruited November 4, 2012)
First Submitted: October 28, 2012 (at clinicaltrials.gov)

Brief Summary

Humeral shaft fractures represent 1-3% of all fractures and 20% of the humeral fractures. These fractures have historically been treated mainly conservatively with good results. Recent development in fracture treatment and findings that certain fracture types are more prone to non-union and bracing-related functional problems of adjacent joints are somewhat common have caused increasing interest in treating these fractures surgically. Return to activities is also considered to be quicker among surgically treated patients.

The purpose of this study is to evaluate effectiveness and cost-effectiveness of surgical treatment of humeral shaft fractures. Patients with a unilateral humeral shaft fracture who are willing to participate in the study after informed consent are randomly assigned to two different treatment methods:

Surgical treatment with an open reduction and internal fixation with a 4,5mm locking plate.
Conservative treatment with functional bracing

The randomization is done using blocked randomization (block sizes are not known by the enrolling or assigning physician) and stratification is done according to fracture type (AO-OTA type A vs. type B/C) and radial nerve status (total/subtotal motor palsy vs. no palsy).

Standard follow-up visits at 6 weeks, 3, 6 and 12 months are arranged. Later follow-up visits are arranged at 2, 5 and 10 years for the study purpose. Patients fill evaluation forms and clinical and radiological assessments are made. The physiotherapist doing objective functional measurements is blinded to treatment method. Both study groups receive physiotherapy after the initial treatment.

Study Design
Study Type: Interventional
Interventional Study Model: Parallel Assignment
Number of Arms: 2
Masking: Single Outcomes Assessor
Allocation: Randomized
Enrollment: 100 [Anticipated]

Arms and Interventions
Active Comparator: Conservative treatment
Conservative treatment with functional brace and physiotherapy.
Device: Conservative treatment
Conservative treatment with functional brace applied after 7 days of initial treatment with prefabricated cork splint.
Physiotherapy
Physiotherapy is arranged to both groups at 3 and 9 wks.

Experimental: Operative treatment

Operative treatment with open reduction and internal fixation with 4,5mm locking compression plate. Physiotherapy at 3 and 9 wks.

Procedure: Operative treatment

Operative treatment with open reduction and internal fixation using 4,5mm locking compression plate.

Physiotherapy

Physiotherapy is arranged to both groups at 3 and 9 wks.

Outcome Measures

Primary Outcome Measures:

1. Pain at rest and in activity, Change in Numerical Rating Scale (NRS) 0-10
   at 6 wks, 3, 6, 12 mo, 2, 5, 10 years
2. Change in The Disabilities of the Arm, Shoulder and Hand Score (DASH)
   at 6 wks, 3, 6, 12 mo, 2, 5, 10 years

Secondary Outcome Measures:

3. Subjective assessment of the function of the upper extremity
   Numerical Rating Scale (NRS) 0-10 Subjective assessment of the function of the upper extremity
4. Constant Score
5. Elbow ROM
6. Health-related quality of life (15D)
7. Complications
   Incidence of re-fracture, reoperation, infection and iatrogenic radial palsy is recorded and compared between study groups.
8. Union
   Time to union, non-union, malunion Union
9. Cost-effectiveness
   Quality-adjusted life years/months measured as a change in 15D tool, pain-NRS and other outcome measures. Cost-effectiveness
10. Subjective assessment of the function of the upper extremity
    Likert Scale 1-7 Subjective assessment of the function of the upper extremity
11. Subjective assessment of the function of the elbow
    Numerical Rating Scale (NRS) 0-10 Subjective assessment of the function of the elbow

Eligibility

Inclusion Criteria:

- Over 18 years old patient who agrees to the consent to participation in this study
- Unilateral dislocated humeral shaft fracture (dislocation over thickness of the bone cortex, fracture below the level of insertion of pectoralis major muscle and 5 cm above the olecranon fossa)
- Randomization can be done within 10 days and operation within 14 days after the initial trauma
- Patient is willing to participate all follow-up visits

Exclusion Criteria:

- Bilateral humeral shaft fracture
- A significant concomitant trauma of the same upper extremity that warrants operative treatment (fracture, tendon injury, soft tissue trauma)
- Other fracture or abdominal/thoracic trauma that warrants operative treatment

- Open fracture
- Pathological fracture
- Multi-trauma patient
- Vascular injury
- Plexus injury
- Previous trauma in the same upper extremity that causes functional deficit
- Trauma or condition that warrants use of walking aid (crutches, wheelchair etc.)
- Disease that affects significantly general condition of the patient
- Significantly impaired ability to co-operate for any reason (substance abuse, mental disorder, dementia)
- Unwilling to accept both treatment methods

1.2. Final registration – Amended Sections Only

The final protocol submitted to Clinical Trials.gov can be accessed at:
https://clinicaltrials.gov/ct2/show/NCT01719887

Enrollment: ~~100 [Anticipated]~~ 82 [Actual]

Outcome Measures

Primary Outcome Measures:

1. ~~Pain at rest and in activity, Change in Numerical Rating Scale (NRS) 0-10 at 6 wks, 3, 6, 12 mo, 2, 5, 10 years~~
2. ~~Change in The Disabilities of the Arm, Shoulder and Hand Score (DASH) at 6 wks, 3, 6, 12 mo, 2, 5, 10 years months~~
1. The Disabilities of the Arm, Shoulder and Hand Score (DASH) at 12 months

Secondary Outcome Measures:

7. Complications
   Incidence of complications (i.e. non-union, malunion, re-fracture, reoperation, infection and iatrogenic radial palsy) is recorded and compared between study groups.
11. The Disabilities of the Arm, Shoulder and Hand Score (DASH)
    at 6 wks, 3, 6 mo, 2, 5, 10 years
12. Pain at rest and in activity, Numerical Rating Scale (NRS) 0-10
    at 6 wks, 3, 6 mo, 12 mo, 2, 5, 10 years
13. Percentage of patients with acceptable symptom state (PASS)

1.3. Summary of Amendments

Primary and secondary outcomes

- Pain at rest and activities downgraded as secondary outcomes
- DASH at 12 months specified as the single primary outcome and other time points downgraded to secondary outcomes

When we registered the trial in ClinicalTrials.gov, our primary outcome measures were the pain at rest and activities at 6 weeks, 3 months, 6 months and 12 months as well as change in DASH at 6 weeks, 3 months, 6 months and 12 months. The secondary outcomes were as listed above in the original protocol. After discussing within the study group about the complexity of having several outcome measures at different time points we first decided to downgrade other time points than 12 months to secondary outcomes (the change was sent to clinicaltrials.gov on January 23,

2013) and later on we made a decision to have only one primary outcome, DASH at 12 months, since this instrument contains also questions regarding pain at rest and at activities. The change was made to clinicaltrials.gov on August 19, 2016.

- Percentage of patients with acceptable symptom state (PASS)

We added this secondary outcome when preparing our protocol publication in the spring 2017 and it was added to clinicaltrials.gov on May 28, 2017. We felt it would add value to our list of secondary outcomes if we define PASS of DASH score in our study population and define which part of the study group has achieved this at different time points.

Enrollment

- Enrollment from 100 [anticipated] to 82 [actual]

When we first registered the study, we reported the enrollment to be 100 patients. We had done the power analysis which showed 35 patients per group and we decided to have 12,5% lost to follow-up reservation. When we sent our study protocol to the ethical board of Helsinki and Uusimaa Hospital District, we put the correct value of 80 patients to the target field. We first registered the enrollment target to 100 patients and after noticing this mistake we made the correction to clinicaltrials.gov on May 28, 2017 when we unified the registered protocol between clinicaltrials.gov and the accepted protocol paper[1]. The number of enrolled patients became 82 since the enrollment took place in two separate units and we were unable to stop the recruiting exactly at 80 patients. After noticing we had achieved the target, we stopped the enrollment on January 2018.

Be it noticed here that all the above noted amendments to the original protocol were made prior to completion of the trial and before doing any data analysis and prior revealing the allocations of the study groups.

2. PROTOCOL

The final study/trial protocol was published in the BMJ Open (Rämö et al[1])

3. STATISTICAL ANALYSIS PLAN – 1-YEAR RESULTS

3.1. Original Statistical Plan

A description of our original statistical analysis plan was published[1] as follows:

The data will be analyzed using IBM SPSS Statistics V.23 or higher. The results will be reported following the Consolidated Standards of Reporting Trials statement.

The baseline characteristics of the participants will be summarized by group, reported as a mean (SD) or median (first quartile, third quartile) for continuous variables, and count (%) for categorical variables.

Primary statistical analyses will be performed using intention-to-treat basis. For the primary analysis, a mixed-effects model (MM) analysis will be performed using the data set without multiple imputation to compare the mean DASH scores. Treatment group and visits will be included as fixed factors and patient as a random factor. The model will include interactions between treatment and visit. Randomization stratification factors and baseline value will be included as covariates. The treatment effect will be quantified with an absolute difference between the groups in the DASH score with the associated 95% CI and p value at 12 months post-randomization.

The MM model will also be used to analyze secondary outcomes where applicable (pain-NRS at rest and during activities, 15D, CS). For categorical response variables, effects will be analyzed by logistic regression analysis with treatment as the fixed-factor covariate. These secondary outcomes will only be supportive, explanatory or hypothesis-generating (or both), which is why multiplicity is not considered to be a problem.

The adverse events of the study arms will be reported descriptively. If the number of events is large enough, an analysis between study arms will be performed.

All scale variables will be tested for normality with the Kolmogorov-Smirnov test. Variance of homogeneity will be tested using Levene's test. We consider a two-sided p value of 0.05 to indicate statistical significance.

We will perform secondary statistical analyses to identify potential effect-modifying and mediating factors. Potential effect-modifying factors to be tested with regression analyses are age, gender, body mass index, physical activity, smoking, level of education, fracture of dominant/non-dominant arm and position of the fracture. The absence of adverse effects and treatment attendance as intended will be analyzed as a potential effect-mediating factor.

We will also perform an on-treatment analysis if there are patients treated with a non-allocated method because patients declined the allocated treatment after the randomization, thus causing crossover in study arms. A medical reason to change treatment method, practically from conservative treatment to ORIF because of non-union or fracture threatening skin integrity in the early phase of treatment, will not be considered as a crossover. However, we will analyze such patients in a separate subgroup.

3.2. Blinded Data Interpretation Protocol

We used blinded data interpretation in analyzing the results of this trial.[2] The blinded data interpretation protocol was published in our protocol paper[1] as follows:

Before accessing the primary outcome data, the Writing Committee will record a 'Background assumptions' statement, which will contain our definition of MID of the outcome measures and a brief summary of the key statistical analysis used in the evaluation of the outcome data. The document will be signed by the members of the Writing Committee and published as an appendix to the primary publication. After this, the Writing Committee will write two interpretations of the trial results on the basis of a blinded review of the primary outcome data (treatment A compared with treatment B), with the assumption that A is the ORIF group and another assuming that A is the conservatively treated group. Decisions regarding the key analyses and presentation format for the primary publication before data analysis will also be decided in a meeting of the Writing Committee. The minutes of this meeting will be recorded as a statement of interpretation document, which will be signed by all members of the Writing Committee before the unsealing of the randomization.

3.3. Statistical Analysis Plan – Amendments

The statistician doing the data analysis is using Stata version 15.1 (StataCorp LLC, Texas, USA) instead of IBM SPSS Statistics. We consider this a minor technical detail which does not affect the interpretation of our results.

Instead of Kolmogorov-Smirnov test for normality and Levene's test for homogeneity, we will use other techniques, e.g., graphical evaluation.

All P values larger than 0.01 are be reported to two decimal places, and those between 0.01 and 0.001 to three decimal places; P values smaller than 0.001 are be reported as P<0.001. We made this amendment since we did not state this in our protocol paper.

Primary analysis – Amendments

The primary comparison on the effectiveness of the treatment will be performed as a between-group comparison using a mixed-model repeated-measures analysis of variance (MMRM ANOVA). In the original analysis plan we used a term 'MM model' but changed the term to 'MMRM ANOVA' as it is more widely used term. We consider this only a terminological issue not affecting the analysis.

Study group and time of assessment (baseline, 6 weeks, 3, 6 and 12 months) were included as fixed factors, patient as a random factor. The model included interactions between study group and time of assessment. Change from baseline was estimated with baseline value as covariate. An unstructured covariance structure will be assumed. If the model cannot be fitted, compound symmetry will be assumed instead. The number of degrees of freedom will be assessed using Satterthwaite's method. The MMRM model will be used to quantify the treatment effect as the absolute difference between the groups in DASH score with the associated 95% confidence interval (CI) and p-value at 12 months post-randomization.

The main publication with primary time point results was published in the JAMA 2020.

4. STATISTICAL ANALYSIS PLAN – 2-YEAR RESULTS

4.1. Background

The primary comparison showed no statistically significant between-group difference in the primary outcome, DASH, at 12 months after randomization. However, an important finding in the preplanned per protocol analysis of the 1-year results showed that the crossover group (patients allocated to bracing but who underwent secondary surgery to promote the healing of the fracture) had a statistically and clinically significant between-group difference in the primary outcome and most of the secondary outcomes compared to the surgery and the bracing group without crossovers. The recovery of these crossover patients after 12 months remained an important study question for the 2-year follow-up. Therefore, we planned the 2-year follow-up analyses to explore whether the crossover group reaches the outcome scores of the early surgery and the bracing group healing without subsequent intervention after the primary time point of 12 months.

4.2. Statistical methods

Our statistical analysis plan at 2-year follow-up will be as follows:

The primary analysis method for this exploratory study will be per protocol analysis with three groups: initial surgery group, successful bracing group with no surgery during 2-year follow-up, and a secondary surgery group who had late surgery to promote the healing of the fracture during 2 years after randomization. In addition, we will carry out intention-to-treat analysis where the patients are analyzed according their randomization groups and an as-treated analysis where the patients are analyzed per latest treatment modality (surgery/nonoperative) at the different follow-up time points. The number of patients in surgery group increased in subsequent follow-up points as patients allocated to functional bracing were operated during the 2 years.

The comparison between the study groups will be performed using a mixed-model repeated-measures analysis of variance. Study group, study site, and time of assessment (baseline, 6 weeks, 3, 6 and 12 months, and 2 years) will be included as fixed factors, patients as random factors. The model includes interactions between study group and time of assessment. Change from baseline will be estimated with baseline value as covariate. The model will be used to quantify the treatment effect as the absolute difference between the groups in the primary outcome (DASH score, mean and 95% confidence interval [CI], and p-value) at 2 years after randomization. A similar model will be used to analyze secondary outcomes where applicable (pain-NRS at rest and during activities, 15D, Constant-Murley Score). For categorical response variables, effects will be analyzed using marginal logistic regression analysis.

Be it reiterated here that the primary comparison for 2-year follow-up (per protocol) will be different from the method of primary comparison at the primary time point of 12 months (intention-to-treat). The rationale behind this is that with this analysis we are primarily exploring whether the patients who underwent late surgery will reach the results of the patients with successful healing of the fracture with initially allocated treatment at 2 years. Because of the potential for type 1 error due to multiple comparisons, all findings for analyses of the 2-year follow-up should be interpreted as exploratory. The statistical model in the analyses allows missing data. No data will be thus imputed. Patients with at least some data can be included in the analyses.

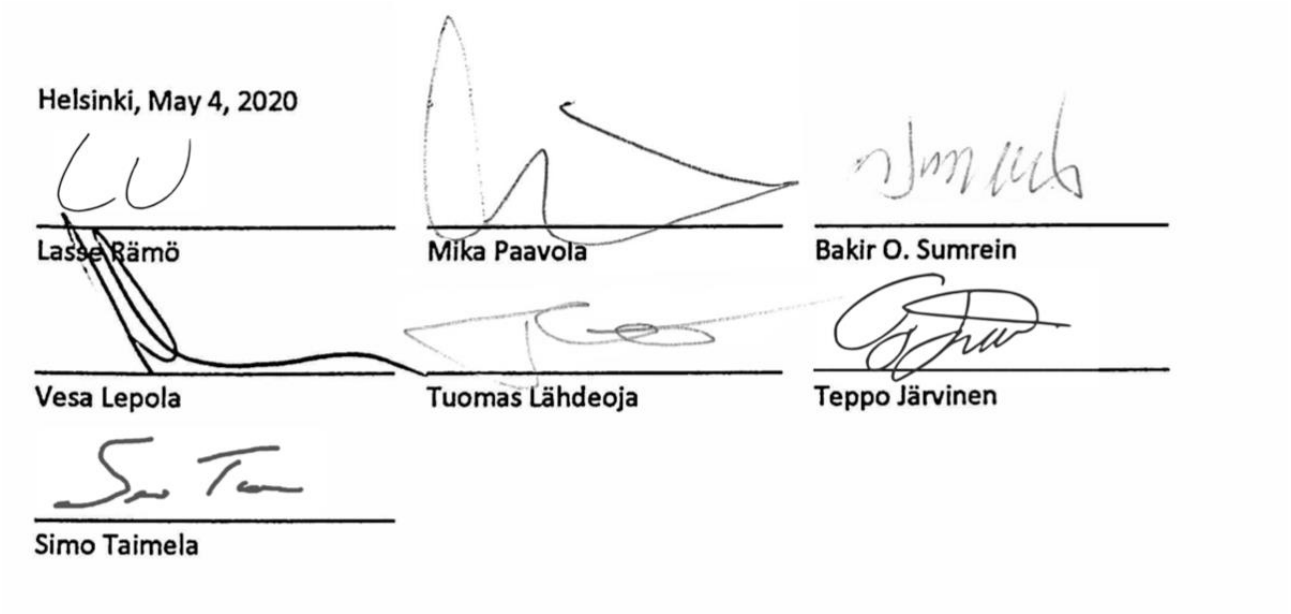
An independent statistician will do all the analyses according to the preplanned statistical analysis plan. The threshold for statistical significance will be set at level 0.05 with 2-sided testing. The data will be analyzed using Stata version 15.1 with the “mixed” procedure (StataCorp LLC, Texas, USA).

4.3. Implementation of Analysis Plan

This SAP will be used as a work description for the statistician performing the analyses. All analyses will be performed by the same statistician and none of the clinical investigators involved in this trial will perform any of the statistical analyses.

Results will be presented to the trial Writing Committee; any uncertainties will be clarified from the statistician. Blinded data interpretation is not used at 2 years as the number of patients in each of the study groups were revealed to the Writing Committee at 1-year.

Be it reiterated here that the statistical analysis plan for 2-year results was decided before having any results from the statistician.



5. STATISTICAL ANALYSIS PLAN – 5-YEAR RESULTS

5.1. Background

The primary comparison showed no statistically significant between-group difference in the primary outcome, DASH, at 12 months after randomization. However, an important finding in the preplanned per protocol analysis of the 1-year results showed that the crossover group (patients allocated to bracing but who underwent secondary surgery to promote the healing of the fracture) had a statistically and clinically significant between-group difference in the primary outcome and most of the secondary outcomes compared to the surgery and the bracing group without crossovers. The recovery of these crossover patients after 12 months remained an important study question for the 2-year follow-up. We found that the results of these patients with secondary surgery remained inferior at two years compared to those with initial surgery and those with successful bracing. Therefore, we planned that at 5-year follow-up we will perform a similar per protocol analysis in addition to primary intention-to-treat analysis.

5.2. Statistical methods

Our statistical analysis plan at 5-year follow-up will be as follows:

The primary analysis method for this study will be intention-to-treat analysis: surgery group and bracing group. In addition, we will carry out a per protocol analysis with three groups: initial surgery group, successful bracing group with no surgery during 5-year follow-up, and a secondary surgery group who had late surgery to promote the healing of the fracture during 5 years after randomization.

The comparison between the study groups will be performed using a mixed-model repeated-measures analysis of variance. Study group, study site, and time of assessment (baseline, 6 weeks, 3, 6 and 12 months, 2 and 5 years) will be included as fixed factors, patients as random factors. The model includes interactions between study group and time of assessment. Change from baseline will be estimated with baseline value as covariate. The model will be used to quantify the treatment effect as the absolute difference between the groups in the primary outcome (DASH score, mean and 95% confidence interval [CI], and p-value) at 5 years after randomization. A similar model will be used to analyze secondary outcomes where applicable (pain-NRS at rest and during activities, 15D, Constant-Murley Score). For categorical response variables, effects will be analyzed using Fisher exact test.

The statistical model in the analyses allows missing data. No data will be thus imputed. Patients with at least some data can be included in the analyses.

An independent statistician will do all the analyses according to the preplanned statistical analysis plan. The threshold for statistical significance will be set at level 0.05 with 2-sided testing. The data will be analyzed using Stata version 15.1 with the “mixed” procedure (StataCorp LLC, Texas, USA).

5.3. Implementation of Analysis Plan

This SAP will be used as a work description for the statistician performing the analyses. All analyses will be performed by the same statistician and none of the clinical investigators involved in this trial will perform any of the statistical analyses.

401 **References**

402

403 1. Rämö L, Taimela S, Lepola V, Malmivaara A, Lähdeoja T, Paavola M. Open reduction and internal fixation of
404 humeral shaft fractures versus conservative treatment with a functional brace: a study protocol of a randomised
405 controlled trial embedded in a cohort. *BMJ Open*. 2017;7(7):e014076. doi:10.1136/bmjopen-2016-014076.

406 2. Järvinen TLN, Sihvonen R, Bhandari M, et al. Blinded interpretation of study results can feasibly and effectively
407 diminish interpretation bias. *Journal of Clinical Epidemiology*. 2014;67(7):769-772.
408 doi:10.1016/j.jclinepi.2013.11.011.

409