
# DISC

## Dupuytren's Interventions Surgery vs. Collagenase

STATISTICAL ANALYSIS PLAN
Draft Version 1.0

**York Trials Unit**
Department of Health Sciences
University of York
York, YO10 5DD

**Version date:** 29/09/2021
**Author(s):** Ada Keding and Charlie Welch
**Chief Investigator:** Prof. Joseph Dias
**Trial Manager**: Dr Puvan Tharmanathan

# Contents

1. Trial objectives ..... 3
2. Design ..... 3
3. Sample size ..... 3
4. Randomisation ..... 4
5. Outcome data ..... 4
5.1 Patient Evaluation Measure ..... 4
5.1.1 PEM - Hand Health Questionnaire (11 items) ..... 5
5.1.2 PEM – Hand Health and Overall Assessment Questionnaire (14 items) ..... 5
5.1.3 PEM – Treatment Questionnaire (5 items) ..... 6
5.2 Unité Rhumatologique des Affections de la Main Scale ..... 6
5.3 Michigan Hand Questionnaire ..... 6
5.4 Single Assessment Numeric Evaluation ..... 7
5.5 Overall hand assessment ..... 7
5.6 Joint measurements ..... 7
5.6.1 Recurrence ..... 8
5.6.2 Range of motion ..... 8
5.7 Complications ..... 8
5.8 Further procedures ..... 8
5.9 Photographs in clinic ..... 8
5.10 Participant photographs ..... 9
5.11 EQ-5D-5L ..... 9
5.12 Resource use ..... 9
6. Other data collected ..... 9
6.1 Treatment preferences ..... 9
6.2 Demographic data ..... 9
6.3 Condition history and diathesis indicators ..... 10
6.4 Comorbidity ..... 10
6.5 Concomitant medications ..... 10
6.6 Condition details and clinical assessment ..... 11
6.7 Adverse Events ..... 11
7. Data management ..... 12
7.2 Data ..... 12
7.2.1 Data capture ..... 12
8. Analysis ..... 13
8.1 Baseline data ..... 13
8.2 Primary outcome analysis ..... 14
8.2.1 PEM completion ..... 14
8.2.2 Unadjusted primary outcome summaries ..... 14
8.2.3 Primary analysis model specification and fitting ..... 15
8.2.4 Model diagnostics ..... 15
8.2.5 Reporting ..... 16

8.3 Further analysis of the primary outcome ..... 16
8.3.1 Missing data ..... 17
8.3.2 Other sensitivity analyses of the primary outcome ..... 19
8.3.3 Departures from randomised treatment ..... 20
8.4 Analysis of secondary effectiveness outcomes ..... 21
8.4.1 PEM Overall Questionnaire (14 items) ..... 21
8.4.2 PEM Treatment Questionnaire (5 items) ..... 21
8.4.3 Unité Rhumatologique des Affections de la Main Scale ..... 21
8.4.4 Michigan Hand Questionnaire ..... 22
8.4.5 Single Assessment Numeric Evaluation Score ..... 23
8.4.6 Overall hand assessment ..... 23
8.4.7 Recurrence ..... 24
8.4.8 Range of movement ..... 25
8.4.9 Active Extension ..... 26
8.4.10 Stiffness ..... 26
8.4.11 Complications ..... 27
8.4.12 Further treatment ..... 29
8.5 Subgroup analyses ..... 30
8.5.1 Treatment preferences ..... 30
8.6 Adverse events ..... 30
8.7 Photography sub-study analysis ..... 30
8.7.1 Primary sub-study analysis ..... 31
8.7.2 Secondary sub study analyses ..... 31
9. SAP amendment log ..... 31
10. Signatures of approval ..... 31
11. References ..... 33
Appendix A – Participant Flow Diagram ..... 34
Appendix B – Data collection schedule ..... 35
Appendix C – MHQ scoring procedure ..... 36
Appendix D – Baseline table ..... 37
Appendix E-1 – PEM Completion ..... 41
Appendix E-2 – Primary Outcome Reporting ..... 42
Appendix F – Secondary outcome measures ..... 46
Appendix F – Tables for Reporting Adverse Events ..... 50

**This document details the proposed statistical analysis of clinical effectiveness and reporting of results, for the Dupuytren's Interventions: Surgery Vs Collagenase (DISC) trial. The analysis of cost-effectiveness data will be detailed in a separate Health Economics Analysis Plan.**

# 1. Trial objectives

The primary objective of the DISC trial is to investigate whether collagenase injection followed by manipulation is non-inferior to limited fasciectomy surgery for the correction of Dupuytren's contracture of the hand. The primary endpoint on which assessment of non-inferiority will be based is the patient reported scores for part two of the Patient Evaluation Measure (PEM) at 12 months post treatment. Secondary objectives are;

- To investigate if remote measurement of extension deficit using photographs is as good as goniometric measurements in clinic to determine recurrence (Photography sub study).

- To investigate whether the correction achieved after Collagenase injection or surgical correction is maintained to 5 years (if justified by findings from the analysis at 1 year and 2 years).

This analysis plan details the analyses that will be implemented to answer the first two objectives. Analysis of longer term follow up data will be specified in a separate plan if a longer term study is conducted

# 2. Design

DISC is a multi-centre, parallel group, individually randomised controlled non-inferiority trial of collagenase injection and manipulation versus limited fasciectomy surgery for the treatment of Dupuytren's contracture. In addition to the main study, DISC contains a qualitative sub study and a photography sub study. This plan outlines the pre-specified analyses of the main trial and the photography sub study. Eligible and consenting patients will be randomised 1:1 to either collagenase injection and manipulation or limited fasciectomy surgery, using a secure randomisation service. Due to the nature of the interventions under investigation, it is not possible to blind clinicians or participants to their treatment allocation. Both treatments should be delivered within 18 weeks of randomisation (as per referral to treatment time), however where possible sites should deliver treatment within 12 weeks of randomisation. Participants will be followed up at 2 and 6 weeks post treatment and 3, 6, 12 and 24 months post treatment. A flow diagram illustrating the flow of participants through the study is presented in Appendix A.

# 3. Sample size

The primary outcome for DISC is the score obtained for the 11 items in part two of the PEM at 12 months post treatment. Previous survey data collected from a representative sample of 880 patients with Dupuytren's Contracture showed the standard deviation of the scores for the 11 items in part 2 of

the PEM at baseline to be 22 points (unpublished). We estimate that a 6 point difference on the PEM at 12 months post-treatment is the threshold at which differences become relevant to the patient, and represents an appropriate non-inferiority margin. An effective sample size of 568 participants (284 per arm) is required to obtain 90% power to assess non-inferiority of collagenase based on the upper limit of a two sided 95% confidence interval (CI) (equivalent to a one-sided 97.5% CI) for the treatment difference ($\delta$ = collagnase – surgery) at 12 months post treatment, assuming a non-inferiority margin of 6 points and a standard deviation of 22 points. This calculation is based upon the sample size needed for 90% power in a one-sided independent samples t-test of size 2.5%, of $H_0: \delta \geq 6$ vs $H_1: \delta < 6$, ignoring informative baseline covariates (reference joint and baseline PEM score). If $\delta = 6$, then we would expect the null hypothesis that collagenase is inferior to surgery to be rejected in 2.5% of hypothetical repetitions of the trial. If $\delta = 0$, then we would expect the null hypothesis that collagenase is inferior to surgery to be (correctly) rejected in 90% of hypothetical repetitions of the trial (assuming an analysis based on a one sided independent samples t-test of size 2.5%). Assuming 20% attrition at the 12 month follow up, the total target sample size is 710.

# 4. Randomisation

The randomisation sequence was designed by the study statistician and is accessed via telephone or the internet using a secure, central randomisation service hosted by Sealed Envelope Ltd. This will ensure adequate allocation concealment for the study. This service will record information to identify all potential participants and their eligibility to avoid inappropriate entry of patients into the trial. The research team at each study site will access the system to complete randomisation following participant consent and completion of baseline assessments. Access to the system for representatives at individual sites will be coordinated and controlled by the trial project team at YTU. Both participants and clinicians will not be blind to allocation. The randomisation system has been designed to allocate participants 1:1 to each of the two study treatment arms. The system uses block randomisation with randomly varying block sizes, stratified by reference joint (MCP or PIP). The actual allocation sequence has been generated using a randomly generated seed provided by a statistician not involved in the ongoing or final analysis of the study.

# 5. Outcome data

Outcome data will be collected using both participant and clinician completed CRFs at 8 different time points. A table summarising the data collection schedule is given in Appendix B.

## 5.1 Patient Evaluation Measure

The PEM [1] is a validated 19 item patient reported outcome measure comprised of three parts.

- Part 1 is comprised of 5 items which ask the participant about their experience of treatment and the care they received.

- Part 2 is comprised of 11 items which ask the participant about their hand's functionality and any pain they may experience.

- Part 3 is comprised of 3 items and provides an overall assessment of the participant's feelings about their hand condition and treatment.

The participant assigns each item an integer score between 1 and 7, where higher scores indicate worse outcomes/experiences. Participants complete the PEM at baseline, immediately prior to treatment delivery and at 3, 6, 12 and 24 months post treatment. Participants are instructed to complete the PEM with respect to the hand designated as the study reference hand. PEM data will be scored as follows.

### 5.1.1 PEM - Hand Health Questionnaire (11 items)

The primary outcome for DISC is the score obtained for the 11 items in Part 2 of the PEM, known as the Hand Health Questionnaire. The Hand Health Questionnaire will be scored as follows:

- Subtract 1 from all responses, so that each item has an integer score between 0 and 6.

- If the participant has fewer than three missing items in part 2, then the missing item(s) are imputed with the mean of the non-missing items. If the participant has more than two missing items then an overall score cannot be calculated and is considered missing.

- Sum the scores for the 11 items to obtain a raw score. This should be between 0 and 66.

- Calculate their percentage disability by generating a normalised score between 0 and 100. A higher score indicates greater disability.

### 5.1.2 PEM – Hand Health and Overall Assessment Questionnaire (14 items)

An overall score for the 14 items in parts 2 and 3 will also be calculated as follows:

- Subtract 1 from all responses, so that each item has an integer score between 0 and 6.

- If the participant has fewer than three missing items in part 2, and no missing items in part 3, then the missing item(s) in part 2 are imputed with the mean of the non-missing items in part 2. If the participant has more than two missing items in part 2, or any missing items in part 3, then an overall score is not calculated and is considered missing.

- Sum the scores for the 14 items to obtain a raw score. This should be between 0 and 84.

- Calculate their percentage disability by generating a normalised score between 0 and 100. A higher score indicates greater disability.

This score will only be calculated and reported for the four post treatment follow up time points. This is because some of the items in part 3 are not applicable to participants prior to treatment delivery.

### 5.1.3 PEM – Treatment Questionnaire (5 items)

An overall score for the 5 items in part 1 will also be calculated as follows:

- Subtract 1 from all responses, so that each item has an integer score between 0 and 6.

- If the participant has one missing item in part 1, then this item is imputed with the mean of the 4 non-missing part 1 items. If the participant has more than one missing item, then an overall score cannot be calculated and is considered missing.

- Sum the scores for the 5 items to obtain the treatment questionnaire score. This should be between 0 and 30.

This score will only be calculated and reported for the four post treatment follow up time points. This is because the items in part 1 are not applicable to participants prior to treatment delivery.

## 5.2 Unité Rhumatologique des Affections de la Main Scale

The URAM is a validated, 9 item, disease specific disability scale assessing the difficulty experienced by the participant in performing particular tasks or movements [2]. Each item is scored between 0 and 5 with higher scores indicating greater difficulty. Participants complete the URAM at baseline and 3, 6, 12 and 24 months post treatment.

- If more than two items are missing then an overall URAM score cannot be calculated and should be considered missing.

- If the participant has fewer than three missing items, then any missing items are imputed with the mean of the non-missing items. If a participant has more than two missing items, then an overall score is not calculated and is considered missing.

- The overall score is calculated by summing the scores for all nine items. The resulting score should be between 0 and 45 with higher scores indicating greater difficulties.

## 5.3 Michigan Hand Questionnaire

The MHQ is a validated, 63 question measure featuring 6 domains: overall hand function, activities of daily living, work performance, pain, aesthetics and patient satisfaction with hand function [3, 4]. Each item is scored between 1 and 5, with some items pertaining to a specific hand and some pertaining to both hands.

The MHQ can be used to generate 7 scale/summary scores for the left hand, right hand and for both hands, giving a total of 21 different summary scores as follows;

- 3x Overall Function scale scores
- 3x Activities of Daily Living scale scores
- 3x Work scale scores
- 3x Pain scale scores
- 3x Aesthetics scale scores
- 3x Satisfaction scale scores
- 3x Overall MHQ scores

Analysis of this data will be concerned with the 6 scale scores and overall MHQ score calculated for the reference hand. Participants complete the MHQ at baseline and 12 and 24 months post treatment. The scoring procedure for the MHQ is given in Appendix C.

## 5.4 Single Assessment Numeric Evaluation

SANE [5] is a single patient reported numerical assessment of hand function, provided by means of a visual analogue scale (VAS). Participants are asked the following question

*How would you rate your hand function today (with normal being 100%)?*

Participants respond by marking a single point on a line between 0 and 100%, to indicate the extent to which they perceive their hand function as being normal. A SANE score is collected at baseline, 2 and 6 weeks post treatment and 3, 6, 12 and 24 months post treatment.

## 5.5 Overall hand assessment

Participants are asked the following question at 3, 6, 12 and 24 months post treatment.

*Overall, how are the problems now, with the hand in which you had treatment, compared to before?*

Participants are given a choice of the following 7 responses; *Cured, Much better, A little better, The same, A little worse, Much worse, Terrible*.

## 5.6 Joint measurements

Goniometric measurements are collected and recorded by investigators in clinic. These will be used to assess range of movement (RoM) and recurrence of contracture post treatment. The collection of this data is summarised in Table 1.

**Table 1:** Joint measurement data collection

| Measurement type | Baseline | Pre treatment | 3 months | 6 months | 12 months | 24 months |
|---|---|---|---|---|---|---|
| Active extension | x | x | x | x | x | x |
| Flexion | x | | x | x | x | x |
| Passive extension | x | x | x | x | x | x |

### 5.6.1 Recurrence

Recurrence at 6 months is defined as a change in extension deficit (measured as passive extension) of the reference joint of 6° or greater between 3 and 6 months post treatment. Recurrence at 12 months is defined as a change in extension deficit (measured as passive extension) of 20° or greater between 3 and 12 months post treatment [6]. This definition will be used to compare recurrence in each arm at 6 and 12 months post treatment.

### 5.6.2 Range of motion

Passive and active extension measurements and flexion measurements will be used to calculate the participant's passive and active RoM at baseline, and 3, 6, 12 and 24 months post treatment and will be used to assess how RoM changes over time.

## 5.7 Complications

Post treatment complications occurring since the previous follow up are recorded at 3, 6, 12 and 24 months post treatment. Details including date of onset, relatedness to study treatment and severity, are recorded as part of the usual trial adverse event reporting.

## 5.8 Further procedures

Data on further treatments and procedures received since the previous follow up will be collected at 3, 6, 12 and 24 months post treatment. This will include the timing of the treatments, whether or not the treatment is related to the reference hand/finger, and details of the treatment received.

## 5.9 Photographs in clinic

Three photographs of participants' hands will be collected at baseline and 3, 6, 12 and 24 months post treatment as follows

- A photograph of the reference hand in extension, taken directly above the hand (anterior view)
- A photograph of the reference hand in extension, taken from the side with the little finger closest to the camera
- A photograph of the reference hand in flexion, taken from the side with the little finger closest to the camera

These images will be used to obtain further measurements of active extension and flexion of the reference digit at each time point. The joint measurements obtained using these images will be used to validate and supplement goniometric measurements and will also be used for secondary analyses in the photography sub study.

## 5.10 Participant photographs

Participants consenting to the photography sub study are shown how to take the required photographs (the same images as taken in clinic) of their hand at baseline and provided with detailed instructions. Sub study participants are asked to take standardised photographs of their study reference hand at baseline (as soon after the baseline visit as possible) and 3, 6, 12 and 24 months post treatment, which are then sent to the study team. Assessing the agreement between the goniometric measurements and measurements obtained using these images will be the primary goal of the photography sub study.

## 5.11 EQ-5D-5L

The EQ-5D-5L is a validated instrument for assessing health related quality of life, comprised of 5 items, with 5 levels of response and a single general health status VAS [7]. Participants complete this questionnaire at baseline, 2 and 6 weeks post-treatment and 3, 6, 12 and 24 months post-treatment. Full details regarding the scoring and analysis of the EQ-5D-5L are given in the Health Economics Analysis plan.

## 5.12 Resource use

Resource use data are collected in both participant and investigator CRFs at 3, 6, 12 and 24 months post treatment for the cost effectiveness analysis. Full details of the resource use data collected and the analysis of these data is given in the Health Economics analysis plan.

# 6. Other data collected

Additional data relating to treatment preferences, comorbidity, concomitant medications, condition history and numerous other baseline characteristics are collected. A table summarising the data collection schedule is given in Appendix B.

## 6.1 Treatment preferences

Participants are asked about their treatment preferences at baseline. Participants indicate whether they would prefer to receive collagenase, limited fasciectomy or have no preference. The baseline CRF is clear that their response to this question will have no impact on their chances of receiving a particular treatment. This data will be used as part of a subgroup analysis to explore possible treatment effect heterogeneity due to receipt (or otherwise) of the preferred treatment.

## 6.2 Demographic data

Participants provide the following demographic details at baseline;

- Date of birth
- Sex
- Ethnicity
- Tobacco smoking status (Never, Current, Previous)
- Alcohol intake (Units per week)
- Which hand is their dominant hand

## 6.3 Condition history and diathesis indicators

At baseline participants provide the following information regarding their history of symptoms and treatment of Dupuytren's disease and the presence of any diathesis indicators;

- Age at which they first experienced Dupuytren's contracture
- Presence of bilateral disease
- Receipt of past treatment with surgery and/or collagenase
- Family history of Dupuytren's contracture
- History of Garrod's pads and details of current symptoms if applicable
- History of Peyronie's disease if applicable
- History of Ledderhose disease and details of current symptoms if applicable

## 6.4 Comorbidity

At baseline participants are asked to provide information about comorbid conditions they may be suffering from including how long they have had the condition, whether it is being treated at present, and whether the treatment provides satisfactory control/relief.

## 6.5 Concomitant medications

Participants are asked to provide the following details at baseline, to ensure they are not contraindicated to receipt of collagenase;

- Use of anticoagulants, for reasons other than a diagnosed coagulation disorder
- Use of anti-platelet agents
- Use of tetracycline antibiotics in previous 14 days

At baseline they are also asked to provide the following details regarding their use of concomitant medications;

- Name of medication
- Reason for use
- Dose
- Frequency
- Route of administration

- Date they started using the medication (if known)
- Date they stopped using the medication (if known and use is not ongoing)

At the four post treatment follow ups they will provide an update about their use of concomitant medication since the previous follow up.

## 6.6 Condition details and clinical assessment

At baseline participants undergo a brief clinical assessment which will, in addition to their goniometric measurements and patient reported assessments, provide an overall summary of the current state of their condition.

- Digits and joints currently affected
- Number of digits/joints affected on the hand/digit being treated as part of the study
- Joint that will be used as the designated study reference digit

## 6.7 Adverse Events

An adverse event (AE) is defined as any untoward medical occurrence experienced by any DISC study participant, which may or may not have a causal relationship with the study treatment. An AE is defined as a serious adverse event (SAE) if it fulfils one, or more, of the following criteria;

- Resulted in death
- Was life threatening
- Resulted in inpatient hospitalisation or prolongation of existing hospitalisation
- Resulted in persistent or significant disability or incapacity
- Resulted in congenital anomaly or birth defect
- Required a surgical or medical intervention to treat/prevent any of the above

All SAE/AEs will be reported and followed up using SAE/AE report forms. The following data are collected;

- Whether or not the AE/SAE was a complication associated with the study treatment
- Specific information about the AE/SAE
- Action taken/treatment given
- Date of onset
- Date of remission (if not ongoing)
- Outcome
- Treatment received
- Relationship to study treatment
- Expectedness
- Seriousness

- Details of death
- Follow up details

# 7. Data management

## 7.2 Data

### 7.2.1 Data capture

Data is principally collected on paper CRFs, completed by both the study participants and investigators. A copy of the CRFs with the variable names from the database (known as 'specs') are kept by the Trial Statistician in the Statistical Master File. The paper CRFs approved for DISC are as follows;

- Confirmation of eligibility CRF (completed during baseline visit)
- Participant baseline CRF
- Investigator baseline CRF
- Supplementary baseline joint measurements CRF
- Participant pre-treatment CRF (completed during treatment delivery visit)
- Treatment delivery CRF
- Participant week 2 CRF
- Participant week 6 CRF
- Participant month 3 CRF
- Investigator month 3 CRF
- Participant month 6 CRF
- Investigator month 6 CRF
- Participant month 12 CRF
- Investigator month 12 CRF
- Participant month 24 CRF
- Investigator month 24 CRF
- Investigator month 3/6/12/24 supplementary outpatient hospital visits CRF
- Investigator month 3/6/12/24 supplementary inpatient hospital visits CRF
- Investigator month 3/6/12/24 supplementary accident and emergency visits CRF
- Supplementary concomitant medication CRF
- Adverse event initial report form
- Serious adverse event initial report form
- Adverse event follow up form

There will also be a set of photographic data comprised of sets of three photographs of the study reference hand taken by the study investigators during clinic visits at the following time points;

- Baseline

- Pre-treatment
- 3 months post treatment
- 6 months post treatment
- 12 months post treatment
- 24 months post treatment

For participants consenting to participation in the photography sub study, there will be an additional five sets of photographs taken by the participant at baseline and at 3, 6, 12 and 24 months post treatment.

# 8. Analysis

All analyses will be conducted following intention to treat principles on a locked dataset using the latest available version of Stata. Continuous data will be summarised in terms of the non-missing sample size, mean, standard deviation, median, interquartile range and range, and categorical data will be summarised in terms of frequencies and proportions. The primary objective of the analysis is to investigate whether there is sufficient evidence to reject the hypothesis that outcomes (as measured by the 11 items in part 2 of the PEM) at 12 months post-treatment among patients randomised to collagenase are inferior to those of patients randomised to surgery, assuming a non-inferiority margin of 6 points. This assessment will be primarily based on whether the upper limit of the two-sided 95% CI for the difference between treatments at 12 months post treatment obtained from the primary analysis model (see Section 8.2) is greater than 6. One of the key secondary objectives is to compare the extent to which the correction achieved following treatment is maintained over the 12 months following treatment. This will be assessed by comparing rates of recurrence at 12 months post-treatment, where recurrence at 12 months is defined as a reduction in passive extension of 20 degrees between 3 and 12 months post-treatment (see Section 5.6.1). We will investigate whether there is sufficient evidence to reject the hypothesis that collagenase is inferior to surgery in terms of rates of recurrence at 12 months post-treatment, assuming a non-inferiority margin of 10%. This assessment will be based on whether the upper limits of the two-sided 95% CIs for the absolute differences in recurrence obtained from the fitted model (see Section 8.4.7) are less than 10%, and associated one-sided hypothesis tests of $H_0: d \geq 10\%$, where $d = \Pr(\text{Recurrence} \mid \text{Collagenase}, X) - \Pr(\text{Recurrence} \mid \text{Surgery}, X)$ for various pre-specified representative covariate patterns $X$.

## 8.1 Baseline data

Baseline data will be summarised descriptively by trial arm and overall, both as randomised and as analysed. The “as randomised” population will include all participants who were randomised, but will exclude any ineligible patients randomised in error (likely to be none or very few patients). The “as analysed” population will include all participants whose data contributes to the primary analysis model (i.e. all participants who provide primary outcome data at one or more post-treatment time point). Brief baseline characteristics and treatment status of just the participants not included in the primary analysis will be presented by treatment group. No formal inferential comparisons of baseline data between

groups will be undertaken. Templates of tables used to summarise baseline data are given in Appendix D.

## 8.2 Primary outcome analysis

The primary outcome for DISC is patient reported disability (0 – 100%) at 12 months post treatment as measured using the 11 items in part 2 of the PEM (the Hand Health Questionnaire). This section describes the analysis of PEM scores, with the scoring procedure described in detail in Section 5.1.1.

### 8.2.1 PEM completion

Participants randomised to surgery will tend to wait substantially longer between baseline and treatment delivery. Hence we expect the data to show some degree of systematic imbalance with respect to time elapsed since baseline. The possible influence of this differential delay has been accounted for by referencing the follow-up data collection to the point of treatment delivery. Any changes in the primary outcome between baseline and treatment delivery due to natural progression of the disease are captured by inclusion of a measurement of the primary outcome at treatment delivery. Between group differences in the timing of treatment and follow up will be explored using tabular and graphical summaries as follows:

- A table of the time elapsed between randomisation and PEM completion at each time point by trial arm.
- Plots of time elapsed between randomisation and PEM completion at treatment delivery
- A table summarising the time elapsed between treatment and PEM completion at each post treatment time point (by trial arm), including the proportion of participants completing the PEM within the windows for completion specified in the protocol (+/- 14 days at 3 and 6 months, +/- 3 months at 12 and 24 months).
- Box plot of time elapsed between treatment and PEM completion at 12 months by trial arm.

We will also report the number/proportion of returned participant questionnaires and the number/ proportion for which a valid score for the primary outcome could be calculated, as well as a summary of the different patterns of missingness and the frequency with which they occur. Templates of tables described in this section are given in Appendix E-1.

### 8.2.2 Unadjusted primary outcome summaries

While the primary analysis will be based on the estimated difference between treatments at 12 months obtained from the primary analysis model (see section 8.2.3), we will also present unadjusted summaries of PEM scores at each time point;

- Table of unadjusted PEM scores at each time point including all participants with valid PEM data at these time points together with unadjusted differences and two-sided Wald 95% CIs
- A plot of unadjusted mean PEM score against categorised time point (baseline, pre-treatment, 3, 6, 12 and 24 months) by treatment group

- A plot of unadjusted difference in mean PEM score against categorised time point (baseline, pre-treatment, 3, 6, 12 and 24 months) together with Wald 95% CIs for the difference between groups

Unadjusted estimates of between group differences in PEM scores will be included in a table with the results of the other analyses of the primary outcome. A template of this table is given in Appendix E-2.

### 8.2.3 Primary analysis model specification and fitting

The model fitted for the primary analysis will include all post treatment PEM scores as outcomes, modelling these using a covariance pattern model. This model will be comprised of three levels with post treatment measurement occasions ($i$ = 1, 2, 3, 4), nested in participants ($j$), nested in centres ($k$). Details of the terms included in the model are given in Table 2.

**Table 2:** Terms included in the primary analysis models. Fixed effect terms for treatment by time interactions will also be included.

| Term | Interpretation | Type | Details |
|---|---|---|---|
| $\text{treat}_{jk}$ | Randomised treatment group (collagenase or surgery) of the $j$th participant at recruitment site $k$ | Binary fixed effect | Indicator variable<br>Surgery = 0, Collagenase = 1 |
| $\text{time}_i$ | $i$th post randomisation time point | Categorical fixed effects | 4 levels ($i$ = 1,…,4)<br>Included in model via 3 dummy indicator variables. |
| $\text{baseline}_{jk}$ | Baseline PEM score of the $j$th participant at recruitment site $k$ (assumed linear relationship with outcome) | Continuous fixed effect | Percentage disability (0 – 100).<br>Missing scores will be imputed with a predicted score via regression of the observed PEM scores on the observed baseline MHQ scores and reference joint (MCP or PIP). If MHQ score is missing then baseline URAM score will be used in an analogous manner. If URAM is also missing then the reference joint strata specific mean of the observed baseline measurements will be used. |
| $\text{joint}_{jk}$ | Designated study reference joint (MCP or PIP) of the $j$th participant at recruitment site $k$ | Binary fixed effect | Indicator variable<br>MCP = 0, PIP = 1 |
| $u_k$ | Random intercept for recruitment site $k$ | Random effect | Normally distributed with mean 0 and variance ${\sigma_k}^2$. Independent across centres |
| $\epsilon_{ijk}$ | Residual error (conditional on $b_k$) of the $i$th measurement of the $j$th participant at recruitment site $k$ | Error term | $\epsilon_{jk}$ follows a multivariate normal distribution with mean 0 and covariance matrix $\Sigma_{jk}$ |

This model will be fitted with restricted maximum likelihood estimation, assuming an unstructured covariance matrix $\Sigma_{jk}$, with degrees of freedom being calculated using the method of Kenward and Roger [8]. As discussed, delay between baseline and treatment delivery differs between the two arms, with participants in the surgery arm generally waiting longer between the two time points. A sensitivity analysis investigating the possible impact of this on the results of the primary analysis is specified below (see Section 8.3.2).

### 8.2.4 Model diagnostics

The fitted primary analysis models will be checked using diagnostic plots based on the standardised conditional residuals. Scedasticity will be assessed by plotting these residuals against the conditional fitted values with a non-parametric locally weighted linear smoothing curve (lowess curve) overlaid.

Normality will be assessed using a quantile-quantile plot. Should these diagnostic analyses suggest serious violation of the model assumptions, then the model described above will be refitted using log-transformed outcome measurements, with a value of 0.1 added prior to transformation to accommodate scores of zero. The fitted model will be used, together with an appropriate transformation, to derive differences in means (at each post treatment time point) on the original scale of measurement, at representative values of the fixed effect covariates in the model (both levels of reference joint, mean of the observed baseline PEM scores and marginalising over the centre random effects).

If serious violation of the model assumptions persists, then the PEM scores at 12 months post-treatment will be modelled in isolation, using a univariate mixed effect linear regression model, adjusting for allocation, reference joint type (the stratification factor) and baseline PEM score as fixed effects, and centre as a random effect. If model assumptions remain violated then an equivalent mixed effect generalised linear model with log link will be fitted. Univariate analyses of the PEM scores at other time points will be conducted in a similar manner.

If model assumptions remain severely violated for the univariate analysis of 12 month PEM scores (using both identity and log links), then a semi-parametric analysis of the untransformed outcome measurements will be conducted using a mixed effect ordinal regression model with logit link. This model will include allocation, reference joint type (the stratification factor) and baseline PEM score as fixed effects, and a random intercept for centre. The adequacy of the logit link will be investigated by assessing the parallelism evident in a plot of the link transformed empirical cumulative distribution function, stratified by randomised group, with a probit link considered if the logit link appears to be severely misspecified. The fitted model will be used to derive estimates of the between group difference in means conditional on representative values of the covariates (both levels of reference joint, mean of the observed baseline PEM scores and marginalising over the centre random effects) together with a two-sided 95% CI based on standard errors obtained via the delta method. Similar analyses will be conducted for PEM scores at other time points.

### 8.2.5 Reporting

Estimates of the adjusted and unadjusted mean difference between treatments will be obtained at each time point with appropriate two-sided 95% CIs. Estimated differences in mean PEM score over time will be plotted, together with 95% CIs. The primary analysis will be principally concerned with the estimated difference between groups at 12 months post-treatment, with estimates at other time points being considered secondary outcomes. All fixed effect parameter estimates from the fitted primary analysis model will be reported together with their standard errors and Wald 95% CIs. Templates of tables used to report the findings of the primary analysis are given in Appendix E-2.

## 8.3 Further analysis of the primary outcome

Further analysis of the primary outcome will have two main objectives. Firstly to investigate the robustness of the results of the primary analysis, and secondly to estimate the complier average causal effect (CACE).

### 8.3.1 Missing data

The model used for the primary analysis assumes that missing outcome data are missing at random (MAR) conditional on the fixed effects and non-missing outcomes included in the model. We will use multiple imputation to relax this MAR assumption by imputing missing values conditional on additional pre and post randomisation variables. We will also conduct analyses to investigate the sensitivity of the results to various systematic departures from MAR.

We will initially explore whether there are any obvious patterns of missingness in relation to outcome or baseline characteristics. Reasons for withdrawal and timing of withdrawal will be summarised to examine the extent to which withdrawal is driven by mechanisms plausibly related to outcome and/or treatment allocation. Unadjusted mean PEM score trajectories will be plotted for participants dropping out at each time point (both in aggregate and by allocation).

Associations between key baseline characteristics and missing outcome data and/or withdrawal prior to treatment will be explored using three sets of Firth logistic regression models [9]. Missing values of continuous baseline covariates will be mean imputed when fitting these models. The following baseline characteristics will be assessed

- Age (continuous)
- Sex (male or female)
- Smoking status (current, previous or never)
- Alcohol use (does drink alcohol or doesn't drink alcohol)
- Previous treatment with either limited fasciectomy or collagenase (previous treatment or no previous treatment)
- Presence of comorbidities (Yes or No)
- Number of digits affected (treated as continuous)
- MHQ score (continuous)
- URAM score (continuous)
- EQ-5D general health VAS score (continuous)

The first set of regressions will be used to identify baseline characteristics associated with missing primary outcome data at any post treatment time point. The second will be used to identify baseline characteristics associated with missing primary outcome data at the primary endpoint (12 months) only. The third set will be used to identify baseline characteristics associated with withdrawal prior to treatment. Any baseline variables that are found to improve model fit (compared with the model with all parameters other than the intercept constrained to zero (not omitted)) in a likelihood ratio test of size 10%, will be added as fixed effects to the primary analysis model, and this model refitted. The results of this analysis will be reported in a similar manner to the reporting of the primary analysis (Appendix E-2).

We will also refit the primary analysis model using a multiply imputed dataset. A total of 250 imputations will be generated using multiple imputation by chained equations. The longitudinal structure of the data

will be accounted for by imputing the data in "wide" format. Further clustering by centre will be ignored for the purposes of the imputation model, (given that between centre heterogeneity is expected to be small and is not the focus of the analysis) but will be allowed for in the substantive analysis models fitted to each of the imputed datasets. Imputation will be carried out separately by randomised group to allow for possible interactions between randomised group and other variables included in the imputation model. Only patients that did not withdraw prior to treatment delivery will have their outcome data imputed, although patients that withdrew prior to treatment will be included in the imputation model. Patients who die during follow up will not have outcome data imputed following death. The variables that will be included in the imputation model are given in Table 3. Any baseline variables that are identified as being associated with missingness of post treatment PEM scores and/or non-receipt of treatment (see above) will also be included. Missing values in baseline variables included in the imputation model will be imputed as necessary (with any values imputed via single imputation methods for other analyses removed).

**Table 3:** Details of variables included in the imputation model

| **Variable** | **Type** | **Details** | **Univariate model (if imputation required)** |
|---|---|---|---|
| Post-treatment PEM scores (four time points) | Continuous | Percentage disability (0 – 100) | Predictive mean matching (10 nearest neighbours) |
| Reference joint | Binary | MCP = 0, PIP = 1 | N/A (complete) |
| Baseline PEM score | Continuous | Percentage disability (0 – 100) | Predictive mean matching (10 nearest neighbours) |
| Treatment delivery PEM score | Continuous | Percentage disability (0 – 100) | Predictive mean matching (10 nearest neighbours) |
| Baseline URAM score | Continuous | Integer score between 0 and 45 | Predictive mean matching (10 nearest neighbours) |
| Post-treatment URAM scores (four time points) | Continuous | Integer score between 0 and 45 | Predictive mean matching (10 nearest neighbours) |
| Baseline MHQ score | Continuous | Overall MHQ score between 0 and 100 | Predictive mean matching (10 nearest neighbours) |
| Post-treatment MHQ score (two time points) | Continuous | Overall MHQ score between 0 and 100 | Predictive mean matching (10 nearest neighbours) |
| Passive extension of reference joint at baseline | Continuous | Mean of the available passive extension measurements taken of the reference joint at baseline | Linear regression |
| Passive extension of reference joint at treatment delivery and 3, 6, 12 and 24 months post-treatment | Continuous | Mean of the available passive extension measurements taken of the reference joint at treatment delivery and 3, 6, 12 and 24 months post-treatment | Linear regression |
| Active extension of reference joint at baseline | Continuous | Mean of the available active extension measurements taken of the reference joint at baseline | Linear regression |
| Active extension of reference joint at treatment delivery and 3, 6, 12 and 24 months post-treatment | Continuous | Mean of the available active extension measurements taken of the reference joint at treatment delivery and 3, 6, 12 and 24 months post-treatment | Linear regression |
| Flexion of reference joint at baseline | Continuous | Mean of the available flexion measurements taken of the reference joint at baseline | Linear regression |
| Flexion of reference joint at treatment delivery and 3, 6, 12 and 24 months post-treatment | Continuous | Mean of the available flexion measurements taken of the reference joint at treatment delivery and 3, 6, 12 and 24 months post-treatment | Linear regression |

| | | | |
|---|---|---|---|
| Crossover | Binary | Did not receive allocated treatment = 0, Received allocated treatment = 1 | Logistic regression (with data augmentation if necessary) |
| Complications | Binary | No complication = 0, Complication = 1 | Logistic regression (with data augmentation if necessary) |
| Further treatment | Binary | No further treatment = 0, Further treatment = 1. Defined as additional treatment to the reference digit with collagenase, limited fasciectomy, dermofasciectomy or percutaneous needle fasciotomy | Logistic regression (with data augmentation if necessary) |

The length of burn in will be determined by examining the trace plots of the chained equations algorithm, with a minimum burn in of 10 iterations being used if this appears to be sufficiently many for the algorithm to reach a stable state for all imputed variables. Imputed datasets will be analysed using the same model used for the primary analysis (without the Kenward-Roger degrees of freedom adjustment) with parameter estimates being combined according to Rubin's rules. The results will be reported in a similar manner to the reporting of the primary analysis results (Appendix E-2)

We will also investigate the sensitivity of results to departures from missing at random by imputing missing primary outcome data under a range of missing not at random scenarios. This will follow a pattern mixture modelling approach as implemented in the user contributed Stata command `rctmiss` [10]. At present this command does not support the covariance pattern model used for the primary analysis. We will therefore only use the outcome data from the 12 month post-treatment time point and use a substantive analysis based on analysis of covariance. Two versions of this analysis will be conducted; one including only patients who did not withdraw prior to treatment, and one including all randomised patients. The substantive analysis models will include allocation, study reference joint and baseline PEM score as predictors, with clustering by centre accounted for using a cluster robust standard error. Imputation will be performed allowing the sensitivity parameter of the pattern mixture model to vary from 0 (mean score among those with missing outcome data is the same as the mean score among those with non-missing outcome data, conditional on the covariates) to +/-10 (mean score among those with missing outcome data is +/-10 points larger/smaller than among those with non-missing outcome data, conditional on the covariates) in each group separately and across both groups. The tabular and graphical summaries used to report the results of these sensitivity analysis are given in Appendix E-2.

### 8.3.2 Other sensitivity analyses of the primary outcome

In addition to the sensitivity analyses investigating the possible influence of missing data, a further three sensitivity analyses will be conducted, investigating the possible influence of differential delay between baseline and treatment delivery in each arm, deviations from the scheduled timing of post-treatment follow ups and an analysis adjusting for additional predictors of disease recurrence. The results of these analyses will be reported in a similar manner to the reporting of the primary analysis (Appendix E-2).

Delay between baseline and treatment delivery differs between the two arms, with participants in the surgery arm generally waiting longer between the two time points. The additional delay in the surgery

arm may lead to greater disease progression by the point of treatment delivery. Hence the baseline measurements in the surgery arm may underestimate severity of disease (at treatment) in comparison to those in the collagenase arm, meaning covariate adjustment using the baseline measurements may lead to a treatment effect estimate which is biased toward zero. Thus the treatment effect estimate from the model adjusting for baseline (as in the primary analysis) may be anti-conservative (in the context of non-inferiority). We will therefore supplement this estimate, with an estimate from a model which adjusts for the PEM measurement taken at treatment delivery (in place of the baseline measurement with all other terms remaining the same as the primary analysis). Any missing treatment delivery PEM scores will be imputed with predicted scores via regression of the observed treatment delivery PEM scores on observed baseline PEM scores and reference joint (MCP or PIP). If the baseline PEM score is also missing then the reference joint strata specific means will be used instead.

To investigate the robustness of the primary analysis to deviations from the scheduled timing of study visits/assessments, the primary analysis models will be refitted using only participants who had PEM scores collected within +/-14 days of the scheduled visit date at 3 and 6 months and within +/-91 days of the scheduled visit date at 12 and 24 months. Any missing baseline PEM scores will be imputed as for the primary analysis.

We will investigate whether the inclusion of additional baseline predictors of recurrence has any impact on the results of the primary analysis. The primary analysis model will be refitted including four additional predictors of recurrence collected at baseline; presence/history of bilateral disease (binary), presence/history of Garrods pads, Peyronie's disease or Ledderhose disease (binary), family history (binary) and age at which the patient first experienced Dupuytren's contracture (linear term). This analysis will be based on complete cases only.

### 8.3.3 Departures from randomised treatment

An exploratory analysis of the primary outcome at 12 months post-treatment, will be conducted to estimate the Complier Average Causal Effect (CACE) (i.e. the average causal effect of collagenase compared with limited fasciectomy surgery in the latent principal strata of "compliers"). Compliance will be defined in terms of whether or not the randomly allocated treatment was received as part of the initial trial treatment delivery visit. Receipt of other trial and non-trial treatments following the initial trial treatment, and before providing outcome data 12 months, will be accepted as part of the treatment strategy.

Brief baseline characteristics of participants who do or do not receive the allocated treatment will be reported by allocation, as will available reasons for receiving a different treatment from the one allocated. A two-stage least squares instrumental variable estimator (assuming randomisation to be an adequate instrument) will be used to estimate the CACE. This will be implemented using Stata's `ivregress` command. Missing baseline PEM scores will be imputed following the same approach as used to impute missing baseline PEM scores for the primary analysis. Missing outcome data will be dealt with using inverse probability weighting. The probability of outcome data being non-missing at 12 months post-treatment will be estimated using a Firth logistic regression model with treatment group,

reference joint, baseline PEM score and baseline covariates associated with missingness identified previously (Section 8.3.1) as fixed effects. Estimated weights will be truncated at their 1st and 99th percentile to avoid extreme weights. The tables used to report brief baseline characteristics by compliance status and the estimate of the CACE are given in Appendix E-2.

## 8.4 Analysis of secondary effectiveness outcomes

Descriptive statistics (including extent of missing data) will be presented for all secondary effectiveness outcomes by trial arm at all time points at which they were collected, with changes in these outcomes, and changes in the estimated difference between groups over time illustrated graphically.

### 8.4.1 PEM Overall Questionnaire (14 items)

Scores will be calculated using participant's responses to the items in parts 2 and 3 of the PEM questionnaire at 3, 6, 12 and 24 months post treatment. These will be presented descriptively by trial arm (complete cases only).

### 8.4.2 PEM Treatment Questionnaire (5 items)

Scores will be calculated using participant's responses to the items in part 1 of the PEM questionnaire at 3, 6, 12 and 24 months post treatment. These will be presented descriptively by trial arm (complete cases only)

### 8.4.3 Unité Rhumatologique des Affections de la Main Scale

Participants complete the URAM at baseline and 3, 6, 12 and 24 months post treatment and their responses used to generate a score between 0 and 45. Descriptive statistics for URAM scores will be reported by treatment group for each time point.

These scores will also be analysed using a similar covariance pattern model as used for the primary analysis, with the estimated differences in means (and two-sided 95% CIs) at each time point being reported. This model will include participants with at least one post-randomisation URAM score available (i.e. those with no post-randomisation URAM measurements will be excluded). Any missing baseline URAM scores will be imputed with predicted values via regression of the observed baseline URAM scores on observed baseline PEM scores and study reference joint (MCP or PIP). If baseline PEM score is also missing then missing baseline URAM scores will be imputed with predicted values via regression on the observed baseline MHQ scores and reference joint. If baseline MHQ score is also missing then missing baseline URAM scores will be imputed with the reference joint strata specific means.

The fitted model will be checked using the same techniques used for the primary analysis. Should any of these diagnostics suggest serious violation of the model assumptions, then the outcome data will be log transformed (with a value of 0.1 added to scores prior to transformation to allow for scores of 0) and the model refitted. If serious violation of the model assumptions persists, then the URAM scores at each post-treatment time point will be modelled in isolation, using a univariate mixed effect linear regression model, adjusting for allocation, reference joint type (the stratification factor) and baseline URAM score

as fixed effects, and centre as a random effect. Mixed effect generalised linear models with log link will be used if model assumptions remain severely violated.

If model assumptions remain severely violated for the univariate analyses of the URAM scores (with both identity and log links), then a semi-parametric analysis of the untransformed outcome measurements will be conducted using a mixed effect ordinal regression model with logit link. This model will include allocation, reference joint type (the stratification factor) and baseline URAM score as fixed effects, with random intercepts for centre. The fitted model will be used to derive estimates of the between group difference in means conditional on representative values of the other covariates (both levels of reference joint, mean of observed baseline URAM scores and marginalising over the random centre effects) together with 95% Wald CIs based on standard errors obtained via the delta method.

### 8.4.4 Michigan Hand Questionnaire

Participants complete the MHQ at baseline and 12 and 24 months post treatment, and their scores used to generate 6 scale scores and an overall score. Only MHQ data relating to the reference hand will be analysed and reported. Descriptive statistics for MHQ scores will be reported by treatment group for each time point.

The overall score for the reference hand at each time point will be analysed using a similar covariance pattern model as used for the primary analysis, with the estimated differences in means (and two-sided 95% CIs) at each time point being reported. This model will include participants with an MHQ score for at least one post-randomisation time point (i.e. those with no post-randomisation MHQ measurements will be excluded). Any missing baseline MHQ scores will be imputed with predicted values via regression of the observed baseline MHQ scores on observed baseline PEM scores and study reference joint (MCP or PIP). If baseline PEM score is also missing then missing baseline MHQ scores will be imputed with predicted values via regression on the observed baseline URAM scores and reference joint. If baseline URAM score is also missing then missing baseline MHQ scores will be imputed with the reference joint strata specific mean.

The fitted model will be checked using the same techniques used for the primary analysis. Should any of these diagnostics suggest serious violation of the model assumptions, then the outcome data will be log transformed (with a value of 0.1 added to scores prior to transformation to allow for scores of 0) and the model refitted. If serious violation of the model assumptions persists, then the MHQ scores at each post-treatment time point will be modelled in isolation, using a univariate mixed effect linear regression models, adjusting for allocation, reference joint type (the stratification factor) and baseline MHQ score as fixed effects, and centre as a random effect. Mixed effect generalised linear models with log link will be used if model assumptions remain severely violated.

If model assumptions remain severely violated for the univariate analysis MHQ scores (with both identity and log links), then a semi-parametric analysis using the untransformed outcome measurements will be conducted using a mixed effect ordinal regression model with logit link. This model will include allocation, reference joint type (the stratification factor) and baseline MHQ score as fixed effects, and

random intercepts for centre. The fitted model will be used to derive estimates of the between group difference in means conditional on representative values of the other covariates (both levels of reference joint, mean of observed baseline MHQ scores and marginalising over the random centre effects) together with Wald 95% CIs based on standard errors obtained via the delta method.

### 8.4.5 Single Assessment Numeric Evaluation Score

Participants are asked to provide a SANE score at baseline, 2 and 6 weeks post treatment and 3, 6, 12 and 24 months post treatment. These scores will be analysed using a similar covariance pattern model as used for the primary analysis, with the estimated differences in means (and two-sided 95% CIs) at each time point being reported. This model will include participants with a SANE score for at least one post-randomisation time point (i.e. those with no post-randomisation SANE measurements will be excluded). Any missing baseline SANE scores will be imputed with predicted values via regression of the observed baseline SANE scores on observed baseline PEM scores and study reference joint (MCP or PIP). If baseline PEM score is also missing then missing baseline SANE scores will be imputed with predicted values via regression on the observed baseline MHQ scores and reference joint. If baseline MHQ score is also missing then missing baseline SANE scores will be imputed with predicted values via regression on the observed baseline URAM scores and reference joint. If baseline URAM score is also missing then missing baseline SANE scores will be imputed with the reference joint strata specific means.

The fitted model will be checked using the same techniques used for the primary analysis. Should any of these diagnostics suggest serious violation of the model assumptions, then the outcome data will be log transformed (with a value of 0.1 added to scores prior to transformation to allow for scores of 0) and the model refitted. If serious violation of the model assumptions persists, then the SANE scores at each post-treatment time point will be modelled in isolation, using a univariate mixed effect linear regression models, adjusting for allocation, reference joint type (the stratification factor) and baseline SANE score as fixed effects, and centre as a random effect. Mixed effect generalised linear models with log link will be used if model assumptions remain severely violated.

If model assumptions remain severely violated for the univariate analysis SANE scores (with both identity and log links), then a semi-parametric analysis of the untransformed outcome measurements will be conducted using a mixed effect ordinal regression model with logit link. This model will include allocation, reference joint type (the stratification factor) and baseline SANE score as fixed effects, with random intercepts for centre. The fitted model will be used to derive estimates of the between group difference in means conditional on representative values of the other covariates (both levels of reference joint, mean of baseline SANE score and marginalising over the random centre effects) together with Wald 95% CIs based on standard errors obtained via the delta method.

### 8.4.6 Overall hand assessment

Participants are asked a global question about the problems they experience with the hand that was treated in comparison to before treatment. Responses are given on a 7 item ordered categorical scale from Terrible to Cured. Responses at each time point will first be summarised descriptively and

graphically by treatment group. Responses to the Overall Hand Assessment at 12 months will be analysed using a mixed effect proportional odds logistic regression model, with randomised group, reference joint and baseline PEM score included as fixed effects and centre as a random effect. This model will include all participants with a non-missing overall hand assessment response at 12 months post-treatment. Missing baseline PEM scores will be imputed using predicted values via regression of the observed baseline PEM scores on observed baseline MHQ scores and reference joint (MCP or PIP). If the MHQ score is also missing then missing baseline PEM scores will be imputed using predicted values via regression of the observed baseline PEM scores on the observed baseline URAM scores and reference joint. If the URAM score is also missing then missing baseline PEM scores will be imputed with the reference joint strata specific mean. The estimated odds ratio for allocation (and two-sided 95% CI) will be reported. This model will also be used to derive the absolute difference between groups in the probability of a response of either “Cured” or “Much better”, at representative values of the fixed effect covariates (mean of observed baseline PEM scores both levels of reference joint and marginalising over the random centre effects), together with Wald 95% CIs based on delta method standard errors.

### 8.4.7 Recurrence

Recurrence will be defined in a binary manner (recurred or not recurred) at 6, 12 and 24 months post-treatment. Recurrence at the 6 month follow up is defined as an increase in passive extension deficit of the reference joint of 6° since the 3 month post treatment follow up. Recurrence at 12 and 24 months is defined as an increase in passive extension deficit of the reference joint of 20° since the 3 month post treatment follow up. These binary outcomes will be modelled using mixed effect logistic regression models adjusting for study reference joint (MCP or PIP), study reference digit, baseline PEM score and baseline passive extension measurement as fixed effects with centre as a random effect. These models will include all participants for whom the relevant (i.e. month 6 or month 12) binary indicator of recurrence can be derived using the available data. If perfect prediction occurs due to empty strata at some covariate combinations, then participants with study reference digit indicated as middle, index or thumb will be pooled so that reference digit becomes a three category covariate (little, ring, other). If the issue persists, then the term for study reference digit will be dropped from the model. If the baseline PEM score and/or baseline passive extension measurement is missing, then they will be imputed using predicted values via regression of the observed values on observed baseline MHQ scores and reference joint (MCP or PIP). If the MHQ score is also missing, then missing baseline PEM scores and/or baseline passive extension measurements will be imputed using predicted values via regression of the observed values on the observed baseline URAM scores and reference joint. If baseline URAM score is also missing then missing baseline PEM scores and/or baseline passive extension measurements will be imputed with the reference joint strata specific means. For both analyses, the estimated odds ratios for allocation (and two-sided 95% CIs) will be reported. This model will also be used to derive the absolute differences in risk of recurrence (together with two-sided 95% CIs) at representative values of the covariates (mean of the continuous covariates, both levels of reference joint, two levels of reference digit (ring or little) and marginalising over the centre random effects). We will also test composite null hypotheses about the absolute difference in incidence of recurrence at 12

and 24 months post-treatment. Specifically we will report one-sided p-values from four Wald tests of $H_0: d \geq 10\%$, where $p < 0.025$ indicates statistical significance, and $d = \Pr(\text{Recurrence} \mid \text{Collagenase}, X) - \Pr(\text{Recurrence} \mid \text{Surgery}, X)$ for the four covariate patterns $X$ outlined above (MCP + Ring, MCP + Little, PIP + Ring, PIP + Little).

Due to the Coronavirus pandemic, a substantial proportion of participants were followed up remotely at 3, 6, 12 and 24 months post-treatment, and did not have goniometric measurements taken. Missingness due to remote follow up, as well as attrition, mean a substantial proportion of the post-treatment goniometric measurement data will be missing. The proportion of missing data at each time point will be summarised by randomised group, as will the available reasons this missing data. The principal analyses of recurrence outlined above (based on the available goniometric measurements) will be repeated using two alternative analysis sets. The first analysis set will supplement the available goniometric measurements with measurements obtained using photographs taken by clinical investigators. If photographs taken by the clinical investigators are not available, then measurements obtained using photographs taken by the participant will be used (note this would include photographs submitted by participants in the photography sub-study, as well as photographs submitted by non-sub-study participants who could not attend clinic visits due to the pandemic). The second analysis will use the multiply imputed dataset(s) generated as part of the analyses specified in Section 8.3.1. Both analysis sets will be analysed using the same models used for the analysis of the available goniometric measurements, with estimates and tests being reported analogously.

### 8.4.8 Range of movement

The passive and active range of movement (RoM) of the joints of the reference digit are measured at baseline and 3, 6, 12 and 24 months post treatment. RoM of joints on the reference digit other than the reference joint will be summarised descriptively and graphically. Active and passive RoM of the reference joint will be analysed using a similar covariance pattern models as used for the primary analysis models, with baseline RoM (active or passive as relevant) and reference digit included in the fixed effect specification and baseline PEM score removed. This model will include all participants with at least one post-treatment measurement of the relevant outcome (active or passive RoM). The estimated between group difference in RoM at each post-treatment follow up will be reported together with two-sided 95% CIs. If baseline range of movement is missing, then it will be imputed using predicted values via regression of the observed measurements on observed baseline MHQ scores and reference joint (MCP or PIP). If MHQ score is also missing, then missing baseline RoM will be imputed with the reference joint strata specific means.

Due to the Coronavirus pandemic, a substantial proportion of participants were followed up remotely at 3, 6, 12 and 24 months post-treatment, and did not have goniometric measurements taken. Missingness due to remote follow up, as well as attrition, mean a substantial proportion of the post-treatment goniometric measurement data will be missing. The proportion of missing data at each time point will be summarised by randomised group, as will the available reasons this missing data. The principal analyses of RoM outlined above (based on the available goniometric measurements) will be repeated

using two alternative analysis sets. The first analysis set will supplement the available goniometric measurements with measurements obtained using photographs taken by clinical investigators. If photographs taken by the clinical investigators are not available, then measurements obtained using photographs taken by the participant will be used (note this would include photographs submitted by participants in the photography sub-study, as well as photographs submitted by non-sub-study participants who could not attend clinic visits due to the pandemic). The second analysis will use the multiply imputed dataset(s) generated as part of the analyses specified in Section 8.3.1. Both analysis sets will be analysed using the same model used to analyse the available goniometric measurements, with the results being reported analogously.

### 8.4.9 Active Extension

The active extension of the joints of the reference digit is measured at baseline and 3, 6, 12 and 24 months post treatment. The active extension of joints on the reference digit other than the reference joint will be summarised descriptively and graphically. Active extension of the reference joint following treatment will be analysed using a similar covariance pattern model as used for the primary analysis, with baseline active extension and reference digit added to the fixed effect specification and baseline PEM score removed. This model will include all participants with at least one post-treatment measurement of active extension. The estimated between group difference in active extension at each post-treatment follow up will be reported together with two-sided 95% CIs. If baseline active extension is missing, then it will be imputed using predicted values via regression of the observed measurements on observed baseline MHQ scores and reference joint (MCP or PIP). If MHQ score is also missing, then missing baseline active extension will be imputed with the reference joint strata specific means.

Due to the Coronavirus pandemic, a substantial proportion of participants were followed up remotely at 3, 6, 12 and 24 months post-treatment, and did not have goniometric measurements taken. Missingness due to remote follow up, as well as attrition, mean a substantial proportion of the post-treatment goniometric measurement data will be missing. The proportion of missing data at each time point will be summarised by randomised group, as will the available reasons this missing data. The principal analyses of active extension outlined above (based on the available goniometric measurements) will be repeated using two alternative analysis sets. The first analysis set will supplement the available goniometric measurements with measurements obtained using photographs taken by clinical investigators. If photographs taken by the clinical investigators are not available, then measurements obtained using photographs taken by the participant will be used (note this would include photographs submitted by participants in the photography sub-study, as well as photographs submitted by non-sub-study participants who could not attend clinic visits due to the pandemic). The second analysis will use the multiply imputed dataset(s) generated as part of the analyses specified in Section 8.3.1. Both analysis sets will be analysed using the same model used to analyse the available goniometric measurements, with the results being reported analogously.

### 8.4.10 Stiffness

The maximal flexion of the joints of the reference digit is measured at baseline and 3, 6, 12 and 24 months post treatment. The flexion measurements of joints on the reference digit other than the

reference joint will be summarised descriptively and graphically. Flexion of the reference joint following treatment will be analysed using a similar covariance pattern model as used for the primary analysis, with baseline flexion measurements and reference digit added to the fixed effect specification and baseline PEM score removed. This model will include all participants with at least one post-treatment measurement of flexion. The estimated between group difference in flexion at each post-treatment follow up will be reported together with two-sided 95% CIs. If baseline flexion is missing, then it will be imputed using predicted values via regression of the observed measurements on observed baseline MHQ scores and reference joint (MCP or PIP). If MHQ score is also missing, then missing baseline flexion will be imputed with the reference joint strata specific means.

Due to the Coronavirus pandemic, a substantial proportion of participants were followed up remotely at 3, 6, 12 and 24 months post-treatment, and did not have goniometric measurements taken. Missingness due to remote follow up, as well as attrition, mean a substantial proportion of the post-treatment goniometric measurement data will be missing. The proportion of missing data at each time point will be summarised by randomised group, as will the available reasons this missing data. The principal analyses of flexion outlined above (based on the available goniometric measurements) will be repeated using two alternative analysis sets. The first analysis set will supplement the available goniometric measurements with measurements obtained using photographs taken by clinical investigators. If photographs taken by the clinical investigators are not available, then measurements obtained using photographs taken by the participant will be used (note this would include photographs submitted by participants in the photography sub-study, as well as photographs submitted by non-sub-study participants who could not attend clinic visits due to the pandemic). The second analysis will use the multiply imputed dataset(s) generated as part of the analyses specified in Section 8.3.1. Both analysis sets will be analysed using the same model used to analyse the available goniometric measurements, with the results being reported analogously.

### 8.4.11 Complications

Data pertaining to complications and adverse events related to treatment are collected, in both the investigator completed follow up CRFs and as part of the adverse event reporting. These will be graded according to their severity using an eight level ordinal classification from no complications up to death (Table 4). Grades will be allocated based on the following domains; symptoms, intervention, recovery duration and residual disability. Allocation of a given complication to a particular grade will initially be performed by two observers working independently of one another. A meeting between the two observers will be used to resolve any conflicts in the initial ranking/classification assigned. The justifications for the assignment of a given complication to a particular level will be recorded and will be available for review.

**Table 4:** Ordinal classification used to rank treatment complications;

| Level | Definition |
|---|---|
| None (0) | No complications identified |
| Technical – not complication (1) | Events related to treatment, but not technically complications (e.g. soft or broken cast, over tight splint) |

| | |
|---|---|
| Very minor (2) | Events related to treatment, but possible to immediately remedy (e.g. removal of retained stich, cast soreness (without pressure sores)) |
| Mild (3) | Events related to treatment, but minor based on symptoms and impact and also reversible |
| Moderate (4) | Events related to treatment that require treatment/action (and/or have ongoing consequences), but do not require hospitalisation or re-operation (e.g. additional therapy and nurse follow up for stiffness, resolving infection) |
| Severe (5) | Events related to treatment that require treatment/action (and/or have ongoing consequences), and require hospitalisation or re-operation and may have lasting consequences |
| Devastating (6) | Events related to treatment that require treatment/action (and/or have significant ongoing consequences), and require hospitalisation or re-operation and have significant lasting consequences (e.g. significant and un-resolving chronic regional pain syndrome) |
| Death (7) | Death (unexpected in DISC and only allocated as a complication if it is directly attributable to the treatment or intervention (e.g. anaphylactic reaction to administered drug leading to death) |

The number and proportion of participants experiencing a complication assigned each level of severity in Table 4 will be reported by randomised group (potentially multiple complications per participant). For further analyses, the highest ranked (i.e. worst) complication experienced will be used.

The highest ranked complication experienced will be cross-tabulated with randomised group. Groups will be compared using a mixed effect proportional odds model (i.e. an ordinal regression model with logit link where the effect of each predictor is constrained to be the same for all levels of the outcome) with fixed effects for allocation (Collagenase or Surgery), reference joint type (MCP or PIP), and reference joint extension deficit at baseline, and a random intercept for study recruitment site. If greater than 10% of cases with non-missing complications data are missing an extension deficit measurement at baseline, then this term will be excluded from the model. If less than or equal to 10% of cases with non-missing complications data are missing an extension deficit measurement at baseline, then these missing baseline measurements will be imputed using the mean of the observed measurements conditional on reference joint type (MCP or PIP), reference digit type (Little vs Ring vs Middle/Index/Thumb) and baseline PEM score. If a participant is also missing a baseline PEM score, then the missing baseline extension deficit measurement will be imputed with the mean of the observed data conditional on reference joint type and reference digit type only. The estimated odds ratio for allocation will be reported, together with a two-sided 95% CI. The fitted model will be used to estimate $\Pr(Y \geq j \mid \text{Intervention}, X) - \Pr(Y \geq j \mid \text{Control}, X)$, (and two-sided 95% CIs based on standard errors obtained via the delta method) for each level $j = 1, \ldots 7$, where *X* denotes representative values of the covariates (both levels of reference joint, mean of the observed baseline PEM scores and two levels of reference digit (ring and little)). This analysis will be repeated for a reduced set of complications that excludes skin tears that occur during joint manipulation, as these events are in many cases an expected part of the procedure rather than a treatment complication. The results for this analysis excluding skin tears will be reported in a similar manner to the reporting of the analysis including skin tears.

A partial proportional odds model will also be fitted relaxing the proportional odds assumption for the effect of allocation (but retaining it for the effects of the other covariates). The random intercept for centre will be dropped and cluster robust standard error will be used instead (currently no software implementations of mixed effect partial proportional odds models). The estimated odds ratios for allocation for each outcome level will be reported together with 95% CIs.

### 8.4.12 Further treatment

At each of the 3, 6, 12 and 24 month post treatment follow up visits, data pertaining to further treatments of the reference digit since the last visit are collected. If participants undergo further collagenase injection, limited fasciectomy surgery, dermofasciectomy or percutaneous needle fasciotomy to the reference digit at any point during follow up, then this will be counted as further treatment. Ongoing hand therapy and/or physiotherapy appointments to treat chronic regional pain syndrome, stiffness, swelling or scar problems, will also be considered further treatment. Participants who undergo more than six outpatient follow up visits for hand therapy and/or physiotherapy will have the details of these appointments reviewed on an individual basis to determine whether this extended duration of therapy should be considered to be further treatment or routine care.

The number of participants receiving further treatment to the reference digit will be reported by allocation, and the unadjusted odds ratio and risk difference will be reported (together with two-sided 95% CIs). The proportion of participants undergoing further treatment to the reference digit will be modelled using a mixed effects logistic regression model. This model will include allocation, study reference joint type (MCP or PIP), study reference digit and baseline PEM score as fixed effects with clustering by centre modelled using a random intercept. If perfect prediction occurs due to empty strata for some covariate combinations, then participants with study reference digit indicated as middle, index or thumb will be pooled so that reference digit becomes a three category covariate. If the issue persists, then the term for study reference joint will be dropped from the model. If baseline PEM score is missing then it will be imputed using predicted values via regression on baseline MHQ score. If MHQ score is also missing then baseline URAM score will be used. If baseline URAM score is also missing then mean imputation will be used. This model will be used to derive adjusted odds ratios and risk differences for allocation (together with two-sided 95% CIs).

We will also explore differences between the two groups in time elapsed between initial treatment and undergoing a further procedure to the reference digit. The proportion of participants yet to undergo a further procedure to the reference digit against time since treatment will be summarised with a Kaplan-Maier curve stratified by treatment group. Time from treatment to the first further procedure to the reference digit will be analysed using Cox proportional hazards regression adjusting for treatment group, study reference joint type (MCP or PIP), study reference digit and PEM score at baseline. A shared frailty for centre will also be included. Participants who reach the final follow up without having had a further procedure to the reference digit will be considered censored on the date of the final follow up. Any participant who withdraws from hospital follow up without having had a further procedure will be considered censored at the time of withdrawal. The proportional hazards assumption will be

evaluated through inspection of Kaplan-Meier curves and log-log plots and tested using Schoenfeld residuals [11, 12]. The hazard ratio for allocation (and two-sided Wald 95% CI) from the fitted model will be reported.

## 8.5 Subgroup analyses

One pre-specified subgroup analysis of the primary outcome will be undertaken.

### 8.5.1 Treatment preferences

Since it is not possible to blind participants to their treatment/allocation and the primary outcome was self-reported, it is plausible that treatment preferences at baseline may influence PEM responses during follow up. The data collected on treatment preferences at baseline will be used to divide participants into three subgroups based on their actual allocation; allocated to preferred treatment, not allocated to preferred treatment and no preference. PEM scores at 12 months post-treatment will be analysed using a univariate linear regression model with fixed effect terms for allocation, reference joint and baseline PEM score and a random intercept for recruitment site, with main effects of preference subgroup (allocated to preferred treatment group, not allocated to preferred treatment group, no treatment preference) and the interaction of this term with allocation being added as fixed effects. This will be compared with the model without this interaction using a likelihood ratio test (both models will be fitted using maximum likelihood). Should this test suggest improvement in model fit at the 10% level, then the model including the preference subgroup main effect and interaction will be used to derive estimates of the treatment effect at 12 months in each preference subgroup together with two sided 95% CIs.

## 8.6 Adverse events

All adverse events will be detailed, as in Appendix G. The total number of SAEs and NSAEs will reported by allocation and overall with further summaries of this data by type of event, relatedness to study treatment and expectedness also given. The proportion of participants experiencing at least one adverse event which was deemed possibly, probably or definitely related to treatment will be reported by allocation together with odds ratios and risk difference for allocation (and two-sided Wald 95% CIs). The mean, median and mode of the number of adverse events per participant will be presented by allocation.

## 8.7 Photography sub-study analysis

Participants consenting to participation in the photography sub study will have three sets of joint measurements at baseline and at 3, 6, 12 and 24 months post treatment; measurements obtained using a goniometer, measurements obtained using photographs taken in clinic, and measurements obtained using photographs taken by participants at home. The primary goal of the photography sub study is to assess the agreement between measurements obtained using a goniometer and measurements obtained using the photographs taken by participants at home, in order to determine whether the two methods of measurement might feasibly be used interchangeably. The photography sub study will also include a number of additional analyses investigating agreement between goniometric measurements and measurements obtained using patient photos following treatment, agreement between goniometric

measurements and measurements obtained using photos taken in clinic, and predictors of agreement (image quality and time elapsed between each type of measurement).

### 8.7.1 Primary sub-study analysis

The primary sub-study analysis will investigate the agreement between the mean of the three measurements obtained using goniometer at baseline, and the measurement obtained using patient photographs at baseline (or patient photographs taken prior to treatment if baseline photos are unavailable). We will calculate separate 95% limits of agreement for the MCP and PIP joints of participants' reference digits. A range of plots will be used to assess the extent to which the assumptions of these analyses are met. If there is evidence of a relationship between the differences of the two measurements and their magnitude, then steps will be taken to account for this relationship using regression methods [13]. The calculation of the limits of agreement assumes that the differences are approximately normally distributed. The extent to which this assumption is met will be assessed by inspection of a normal quantile-quantile plot. Should there be extreme departures from normality then a non-parametric approach will be used as described in [13]. The differences between the two types of measurement will be plotted against their average, with the estimated limits of agreement overlaid. The extent to which agreement varies by digit will be investigated by generating separate limits of agreement and plots by digit (little, ring or middle/index/thumb). If feasible, the extent to which agreement varies according to image quality will be investigated by generating separate limits of agreement and plots by image quality category. The extent to which time between measurements explains variation in agreement will be explored by regressing the observed differences on the magnitude of the measurements (i.e. the average of the measurements obtained using each method) and the time elapsed between the two measurements.

### 8.7.2 Secondary sub study analyses

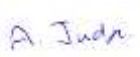
Secondary analyses of the photography sub study will use similar techniques as described for the primary sub study analysis to assess the agreement between goniometric measurements and measurements obtained using participant photographs at the 3 and 6 month post treatment time points. Similar techniques will again be used to assess the agreement between baseline goniometric measurements and measurements obtained using photographs taken in clinic at baseline.

## 9. SAP amendment log

| **Amendment/addition to SAP and reason for change** | **New version number, name and date** |
|---|---|
| | |
| | |
| | |

## 10. Signatures of approval

Sign-off of the final approved version of the Statistical Analysis Plan by the principle investigator and trial statistician(s) (can also include Trial Manager/Co-ordinator)

| **Name** | **Trial Role** | **Signature** | **Date** |
|---|---|---|---|
| Prof. Andrew Judge | Chair of DMC |  | 29/09/2021 |
| Prof. Wendy Baird | Chair of TSC |  | 19/10/21 |
| Prof. Joseph Dias | Chief investigator |  | 01/10/21 |
| Dr. Puvan Tharmanathan | Trial manager |  | 29/09/2021 |
| Mr. Charlie Welch | Trial statistician |  | 29/09/2021 |

# 11. References

[1] Dias J.J., Bhowal B., Wildin C.J., Thompson J.R., Assessing the outcome of disorders of the hand, *Bone & Joint Journal*, 2001; 83(2): 235-40.

[2] Beaudreuil J., Allard A., Zerkak D., Gerber R.A., Cappelleri J.C., et al., Unité Rhumatologique des Affections de la Main (URAM) scale: Development and validation of a tool to assess Dupuytren's disease–specific disability, *Arthritis Care & Research*, 2011; 63(10): 1448-1455.

[3] Chung K.C., Pillsbury M.S., Walters M.R., Hayward R.A., Reliability and validity testing of the Michigan Hand Outcomes Questionnaire, *The Journal of Hand Surgery*. 1998; 23(4): 575-587.

[4] Chung K.C., Hamill J.B., Walters M.R., Hayward R.A., The Michigan Hand Outcomes Questionnaire (MHQ): assessment of responsiveness to clinical change. *Annals of Plastic Surgery*, 1999; 42(6): 619-622.

[5] Williams G.N., Gangel T.J., Arciero R.A., Uhorchak J.M., Taylor D.C., Comparison of the single assessment numeric evaluation method and two shoulder rating scales - outcome measures after shoulder surgery, *The American Journal of Sports Medicine*, 1999; 27(2): 214 - 221.

[6] Felici N., Marcoccio I., Giunta R., Haerle M., Leclercq C., Pajardi G., et al., Dupuytren's contracture recurrence project: reaching consensus on a definition of recurrence, *Hand Chirurgie Mikro Chirurgie Plastische Chirurgie*, 2014; 46(6): 350-354.

[7] The EuroQol Group, EuroQol-a new facility for the measurement of health-related quality of life, *Health Policy*, 1990; 16(3): 199-208.

[8] Kenward, M.G., and J.H. Roger.. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, 1997; 53: 983-997

[9] Firth D., Bias reduction of maximum likelihood estimates, *Biometrika*, 1993; 80(1) 27-38

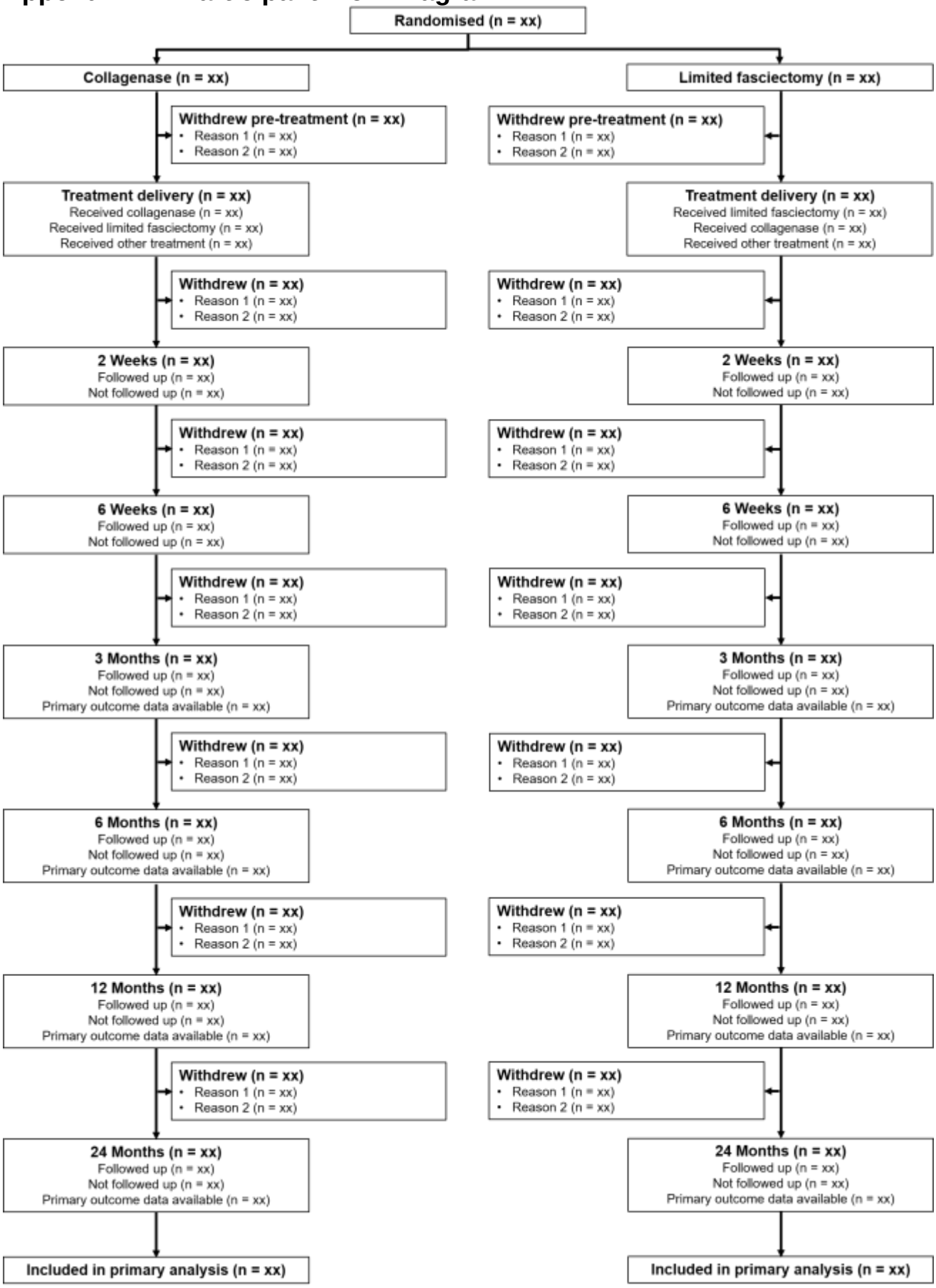
[10] White I.R., rctmiss*:* Stata module to analyse a randomised controlled trial (RCT) allowing for informatively missing data, *Statistical Software Components S458304*, Boston College Department of Economics

[11] Schoenfeld D., Chi-squared goodness of fit tests for the proportional hazards regression model, *Biometrika*, 1980; 67(1): 145-153

[12] Grambsch P.M., Therneau T.M., Proportional hazards tests and diagnostics based on weighted residuals, *Biometrika*, 1994; 81(3): 515-526

[13] Bland J.M., Altman D.G., Measuring agreement in method comparison studies, *Statistical Methods in Medical Research*, 1999; 8: 135-160

# Appendix A – Participant Flow Diagram



# Appendix B – Data collection schedule

Data will be collected using participant and investigator CRFs

| Data Collected | Baseline | Treatment Delivery | 2 Weeks Post Treatment | 6 Weeks Post Treatment | 3 Months Post Treatment | 6 Months Post Treatment | 12 Month Post Treatment | 24 Months Post Treatment |
|---|---|---|---|---|---|---|---|---|
| **Patient Reported** | | | | | | | | |
| PEM | X | X | | | X | X | X | X |
| URAM | X | | | | X | X | X | X |
| MHQ | X | | | | | | X | X |
| EQ-5D-5L | X | | X | X | X | X | X | X |
| SANE | X | | X | X | X | X | X | X |
| Treatment Preference | X | | | | | | | |
| Overall Hand Assessment | | | | | X | X | X | X |
| Resource Use | | | | | X | X | X | X |
| **Demographics and Medical History** | | | | | | | | |
| Demographics | X | | | | | | | |
| Comorbidity | X | | | | | | | |
| Concomitant Medications | X | | | | | | | |
| **Condition Details and Clinical Assessment** | | | | | | | | |
| Condition History | X | | | | | | | |
| Diathesis Indicators | X | | | | | | | |
| Clinical Assessment | X | | | | | | | |
| **Photography** | | | | | | | | |
| Clinician Photographs | X | X | | | X | X | X | X |
| Participant Photographs | Y | Y | | | Y | Y | Y | Y |
| **Treatment Pathway** | | | | | | | | |
| Randomisation | X | | | | | | | |
| Treatment Delivered | | X | | | X | | | |
| **Clinical Outcomes** | | | | | | | | |
| Joint Measurements | X | X | | | X | X | X | X |
| Complications | | X | | | X | X | X | X |
| Further Procedures | | | | | X | X | X | X |
| **Safety Data** | | | | | | | | |
| Contraindications | X | X | | | | | | |
| Adverse Events | | Y | Y | Y | Y | Y | Y | Y |

**X = Mandatory data collection, Y = Conditional/triggered data collection**

# Appendix C – MHQ scoring procedure

- The second item in the pain scale for each hand (MHQ34 and MHQ39) is reversed scored so that 1 = 5, 2 = 4, 4 = 2, 5 = 1
- The second item in the aesthetics scale for each hand (MHQ43 and MHQ47) is reverse scored so that 1 = 5, 2 = 4, 4 = 2, 5 = 1.
- If 50% or more of the items in a given scale are missing then a score for this scale cannot be validly calculated and should be considered missing.
- If strictly less than 50% of the items in a given scale are missing then these should be imputed with the mean of the non-missing items in that scale.
- A raw score for each scale is obtained by summing the responses in that scale.
- The raw scores are normalised to a range 0 - 100 following the procedure described in the table below.
- For the activities of daily living scale, the final scale score is the average of the one handed and two handed scores. For all other scales the final scale score is just the normalised score.

| Scale | Items | Raw Score* ($x$) | Normalisation** | Scale Score*** |
|---|---|---|---|---|
| **Overall Hand Function** | | | | |
| Right | 1, 2, 3, 4, 5 | $5 \leq x \leq 25$ | $y = 5 \times (25 - x)$ | $y$ |
| Left | 6, 7, 8, 9, 10 | $5 \leq x \leq 25$ | $y = 5 \times (25 - x)$ | |
| **Activities of Daily Living** | | | | |
| Right | 11, 12, 13, 14, 15 | $5 \leq x \leq 25$ | $y = 5 \times (25 - x)$ | $\frac{y+z}{2}$ |
| Left | 16, 17, 18, 19, 20 | $5 \leq x \leq 25$ | $y = 5 \times (25 - x)$ | |
| Both | 21, 22, 23, 24, 25, 26, 27 | $7 \leq x \leq 35$ | $z = \frac{25}{7} \times (35 - x)$ | |
| **Work** | | | | |
| Both | 28, 29, 30, 31, 32 | $5 \leq x \leq 25$ | $y = 5 \times (x - 5)$ | $y$ |
| **Pain** | | | | |
| Right | 33, 34, 35, 36, 37 | $5 \leq x \leq 25$ | $y = 0$ if item $33 = 5$<br>$y = 5 \times (25 - x)$ otherwise | $y$ |
| Left | 38, 39, 40, 41, 42 | $5 \leq x \leq 25$ | $y = 0$ if item $38 = 5$<br>$y = 5 \times (25 - x)$ otherwise | |
| **Aesthetics** | | | | |
| Right | 43, 44, 45, 46 | $4 \leq x \leq 20$ | $y = \frac{25}{4}(x - 4)$ | $y$ |
| Left | 47, 48, 49, 50 | $4 \leq x \leq 20$ | $y = \frac{25}{4}(x - 4)$ | |
| **Satisfaction** | | | | |
| Right | 51, 52, 53, 54, 55, 56 | $6 \leq x \leq 30$ | $y = \frac{25}{6}(30 - x)$ | $y$ |
| Left | 57, 58, 59, 60, 61, 62 | $6 \leq x \leq 30$ | $y = \frac{25}{6}(30 - x)$ | |

* For pain scale a higher score indicates more pain. For the other five scales a higher score indicates better hand performance.
** A total MHQ score is obtained by taking the mean of the six scale scores, after reversing the score for the pain scale. The pain scale score is reversed by subtracting this score form 100.
*** The total MHQ score will be considered missing if any of the six scale scores are missing.

# Appendix D – Baseline table

**Table xx**: Baseline characteristics of DISC trial participants by allocation.

| | **As Randomised** (All randomised participants n = ) | | **As Analysed** (Participants included in primary analysis, n = ) | |
|---|---|---|---|---|
| | **Collagenase** (n = ) | **Surgery** (n = ) | **Collagenase** (n = ) | **Surgery** (n = ) |
| **Age (years)** | | | | |
| N | | | | |
| Mean (SD) | | | | |
| Median (Q1, Q3) | | | | |
| Min, Max | | | | |
| **Sex** | | | | |
| Male | | | | |
| Female | | | | |
| Missing | | | | |
| **Ethnicity** | | | | |
| White | | | | |
| Mixed Race | | | | |
| Asian/Asian British | | | | |
| Black/Black British | | | | |
| Chinese or Other | | | | |
| Missing | | | | |
| **Smoking** | | | | |
| Never | | | | |
| Current | | | | |
| Previous | | | | |
| Missing | | | | |
| **Alcohol** | | | | |
| Yes | | | | |
| No | | | | |
| Missing | | | | |
| **Affected hand(s)** | | | | |
| Left | | | | |
| Right | | | | |
| Both | | | | |
| **Study reference hand** | | | | |
| Left | | | | |
| Right | | | | |
| **Study reference digit** | | | | |
| Thumb | | | | |
| Index | | | | |
| Middle | | | | |
| Ring | | | | |
| Little | | | | |
| **Study reference joint** | | | | |
| MCP | | | | |
| PIP | | | | |
| **Number of digits affected** | | | | |
| N | | | | |
| Mean (SD) | | | | |
| Median (Q1, Q3) | | | | |
| Min, Max | | | | |
| **Number of digits affected on reference hand** | | | | |
| N | | | | |
| Mean (SD) | | | | |
| Median (Q1, Q3) | | | | |
| Min, Max | | | | |
| **Number of joints currently affected** | | | | |
| N | | | | |
| Mean (SD) | | | | |
| Median (Q1, Q3) | | | | |

| | **As Randomised** (All randomised participants n = ) | | **As Analysed** (Participants included in primary analysis, n = ) | |
|---|---|---|---|---|
| | **Collagenase** (n = ) | **Surgery** (n = ) | **Collagenase** (n = ) | **Surgery** (n = ) |
| Min, Max | | | | |
| **Number of joints currently affected on reference hand** | | | | |
| N | | | | |
| Mean (SD) | | | | |
| Median (Q1, Q3) | | | | |
| Min, Max | | | | |
| **Age of onset** | | | | |
| N | | | | |
| Mean (SD) | | | | |
| Median (Q1, Q3) | | | | |
| Min, Max | | | | |
| **Previous surgery** | | | | |
| Yes | | | | |
| No | | | | |
| Missing | | | | |
| **Previous CCH injections** | | | | |
| Yes | | | | |
| No | | | | |
| Missing | | | | |
| **Family history of Dupuytren's disease** | | | | |
| Yes | | | | |
| No | | | | |
| Missing | | | | |
| **History of Garrod's pads** | | | | |
| Yes | | | | |
| No | | | | |
| Missing | | | | |
| **Peyronie's disease** | | | | |
| Yes | | | | |
| No | | | | |
| Not applicable | | | | |
| Missing | | | | |
| **Ledderhose disease** | | | | |
| Yes | | | | |
| No | | | | |
| Missing | | | | |
| **Presence of Comorbidities** | | | | |
| Yes | | | | |
| No | | | | |
| Missing | | | | |
| **Concomitant Medications** | | | | |
| Yes | | | | |
| No | | | | |
| Missing | | | | |
| **Total number of medications used** | | | | |
| N | | | | |
| Mean (SD) | | | | |
| Median (Q1, Q3) | | | | |
| Min, Max | | | | |
| **Passive Extension – MCP** | | | | |
| N | | | | |
| Mean (SD) | | | | |
| Median (Q1, Q3) | | | | |
| Min, Max | | | | |
| **Passive Extension – PIP** | | | | |
| N | | | | |
| Mean (SD) | | | | |

| | **As Randomised** (All randomised participants n = ) | | **As Analysed** (Participants included in primary analysis, n = ) | |
|---|---|---|---|---|
| | **Collagenase** (n = ) | **Surgery** (n = ) | **Collagenase** (n = ) | **Surgery** (n = ) |
| Median (Q1, Q3) | | | | |
| Min, Max | | | | |
| **Passive Extension – DIP** | | | | |
| N | | | | |
| Mean (SD) | | | | |
| Median (Q1, Q3) | | | | |
| Min, Max | | | | |
| **Active Extension – MCP** | | | | |
| N | | | | |
| Mean (SD) | | | | |
| Median (Q1, Q3) | | | | |
| Min, Max | | | | |
| **Active Extension – PIP** | | | | |
| N | | | | |
| Mean (SD) | | | | |
| Median (Q1, Q3) | | | | |
| Min, Max | | | | |
| **Active Extension – DIP** | | | | |
| N | | | | |
| Mean (SD) | | | | |
| Median (Q1, Q3) | | | | |
| Min, Max | | | | |
| **Flexion – MCP** | | | | |
| N | | | | |
| Mean (SD) | | | | |
| Median (Q1, Q3) | | | | |
| Min, Max | | | | |
| **Flexion – PIP** | | | | |
| N | | | | |
| Mean (SD) | | | | |
| Median (Q1, Q3) | | | | |
| Min, Max | | | | |
| **Flexion – DIP** | | | | |
| N | | | | |
| Mean (SD) | | | | |
| Median (Q1, Q3) | | | | |
| Min, Max | | | | |
| **Patient Evaluation Measure** | | | | |
| N | | | | |
| Mean (SD) | | | | |
| Median (Q1, Q3) | | | | |
| Min, Max | | | | |
| **Unite Rhumatologique des Affections de la Main** | | | | |
| N | | | | |
| Mean (SD) | | | | |
| Median (Q1, Q3) | | | | |
| Min, Max | | | | |
| **Michigan Hand Questionnaire** | | | | |
| N | | | | |
| Mean (SD) | | | | |
| Median (Q1, Q3) | | | | |
| Min, Max | | | | |
| **SANE** | | | | |
| N | | | | |
| Mean (SD) | | | | |
| Median (Q1, Q3) | | | | |
| Min, Max | | | | |

| | **As Randomised** (All randomised participants n = ) | | **As Analysed** (Participants included in primary analysis, n = ) | |
|---|---|---|---|---|
| | **Collagenase** (n = ) | **Surgery** (n = ) | **Collagenase** (n = ) | **Surgery** (n = ) |
| **Treatment Preferences** | | | | |
| Collagenase | | | | |
| Surgery | | | | |
| No Preference | | | | |
| Missing | | | | |

**Table xx**: Brief baseline characteristics of participants not included in the primary analysis model.

| | **Collagenase (N = )** | **Surgery (N = )** | **Total (N = )** |
|---|---|---|---|
| **Age** | | | |
| N | | | |
| Mean (SD) | | | |
| Median (Q1, Q3) | | | |
| Min, Max | | | |
| **Sex** | | | |
| Male | | | |
| Female | | | |
| Missing | | | |
| **Study reference joint** | | | |
| MCP | | | |
| PIP | | | |
| **Presence of comorbidities** | | | |
| Yes | | | |
| No | | | |
| Missing | | | |
| **Number of medications used** | | | |
| N | | | |
| Mean (SD) | | | |
| Median (Q1, Q3) | | | |
| Min, Max | | | |
| **PEM** | | | |
| N | | | |
| Mean (SD) | | | |
| Median (Q1, Q3) | | | |
| Min, Max | | | |
| **URAM** | | | |
| N | | | |
| Mean (SD) | | | |
| Median (Q1, Q3) | | | |
| Min, Max | | | |
| **MHQ** | | | |
| N | | | |
| Mean (SD) | | | |
| Median (Q1, Q3) | | | |
| Min, Max | | | |
| **Treatment Preferences** | | | |
| Collagenase | | | |
| Surgery | | | |
| No preference | | | |
| Missing | | | |
| **Treatment received** | | | |
| Collagenase | | | |
| Surgery | | | |
| Other | | | |
| None | | | |
| Missing | | | |

# Appendix E-1 – PEM Completion

**Table xx**: Time elapsed (days) between baseline and PEM completion time points.

| | Collagenase (N = xx) | Surgery (N = xx) | Total (N = xx) |
|---|---|---|---|
| **Pre-Treatment** | | | |
| N (%) | | | |
| Mean (SD) | | | |
| Median (Q1, Q3) | | | |
| Min, Max | | | |
| **3 Months Post Treatment** | | | |
| N (%) | | | |
| Mean (SD) | | | |
| Median (Q1, Q3) | | | |
| Min, Max | | | |
| **6 Months Post Treatment** | | | |
| N (%) | | | |
| Mean (SD) | | | |
| Median (Q1, Q3) | | | |
| Min, Max | | | |
| **12 Months Post Treatment** | | | |
| N (%) | | | |
| Mean (SD) | | | |
| Median (Q1, Q3) | | | |
| Min, Max | | | |
| **24 Months Post Treatment** | | | |
| N (%) | | | |
| Mean (SD) | | | |
| Median (Q1, Q3) | | | |
| Min, Max | | | |

**Table xx**: Time elapsed (days) between treatment delivery and PEM completion at each time point (N* denotes the number of participants completing the PEM within the specified windows for completion at each time point given in the protocol)

| Time Point | Collagenase (N = xx) | Surgery (N = xx) | Total (N = xx) |
|---|---|---|---|
| **3 Months Post Treatment** | | | |
| N (%) | | | |
| N* (%) | | | |
| Mean (SD) | | | |
| Median (Q1, Q3) | | | |
| Min, Max | | | |
| **6 Months Post Treatment** | | | |
| N (%) | | | |
| N* (%) | | | |
| Mean (SD) | | | |
| Median (Q1, Q3) | | | |
| Min, Max | | | |
| **12 Months Post Treatment** | | | |
| N (%) | | | |
| N* (%) | | | |
| Mean (SD) | | | |
| Median (Q1, Q3) | | | |
| Min, Max | | | |
| **24 Months Post Treatment** | | | |
| N (%) | | | |
| N* (%) | | | |
| Mean (SD) | | | |
| Median (Q1, Q3) | | | |
| Min, Max | | | |

# Appendix E-2 – Primary Outcome Reporting

**Table xx:** Descriptive summaries of PEM scores over time

| | Collagenase (N = ) | Surgery (N = ) | Total (N = ) |
|---|---|---|---|
| **Baseline** | | | |
| N | | | |
| Mean (SD) | | | |
| Median (Q1, Q3) | | | |
| Min, Max | | | |
| **Pre-treatment** | | | |
| N | | | |
| Mean (SD) | | | |
| Median (Q1, Q3) | | | |
| Min, Max | | | |
| **Month 3** | | | |
| N | | | |
| Mean (SD) | | | |
| Median (Q1, Q3) | | | |
| Min, Max | | | |
| **Month 6** | | | |
| N | | | |
| Mean (SD) | | | |
| Median (Q1, Q3) | | | |
| Min, Max | | | |
| **Month 12** | | | |
| N | | | |
| Mean (SD) | | | |
| Median (Q1, Q3) | | | |
| Min, Max | | | |
| **Month 24** | | | |
| N | | | |
| Mean (SD) | | | |
| Median (Q1, Q3) | | | |
| Min, Max | | | |

**Table xx:** Estimated differences in mean PEM scores over time for each of the planned analyses of the primary outcome (except the analyses exploring sensitivity to outcome data being missing not at random)

| | Estimated difference (95% CI) |
|---|---|
| **Unadjusted analysis** | |
| Baseline | - |
| Pre-Treatment | |
| 3 Months | |
| 6 Months | |
| 12 Months | |
| 24 Months | |
| **Primary analysis** | |
| 3 Months | |
| 6 Months | |
| **12 Months (primary endpoint)** | |
| 24 Months | |
| **Sensitivity analysis including baseline predictors of missingness** | |
| 3 Months | |
| 6 Months | |
| 12 Months | |
| 24 Months | |
| **Sensitivity analysis using multiply imputed data** | |
| 3 Months | |
| 6 Months | |
| 12 Months | |

| | |
|---|---|
| 24 Months | |
| **Sensitivity analysis adjusting for treatment delivery measurement** | |
| 3 Months | |
| 6 Months | |
| 12 Months | |
| 24 Months | |
| **Sensitivity analysis using data collected within planned follow up window*** | |
| 3 Months | |
| 6 Months | |
| 12 Months | |
| 24 Months | |
| **Sensitivity analysis adjusting for diathesis** | |
| 3 Months | |
| 6 Months | |
| 12 Months | |
| 24 Months | |
| **Complier Averaged Causal Effect Analysis** | |
| 12 Months | |
| **Subgroup analysis** | |
| Received preferred treatment | |
| Did not receive preferred treatment | |
| No treatment preference | |

*The specified timeframes are as follows. Pre-treatment PEM completed within 18 weeks of randomisation, month 3 and month 6 PEM completed within +/-14 days of 3 and 6 months post treatment, month 12 and month 24 PEM completed within +/- 3 months of 12 and 24 months post treatment.

**Table xx:** Sensitivity analysis exploring departures from MAR.

| **Δ (sensitivity parameter)** | **Imputing assuming $\Delta = \delta$ in which arm?** | **Estimated Difference (95% CI)** |
|---|---|---|
| $\delta = 0$ | Collagenase arm only | |
| | Surgery arm only | |
| | Both arms | |
| $\delta = 1$ | Collagenase arm only | |
| | Surgery arm only | |
| | Both arms | |
| $\delta = 2$ | Collagenase arm only | |
| | Surgery arm only | |
| | Both arms | |
| $\delta = 3$ | Collagenase arm only | |
| | Surgery arm only | |
| | Both arms | |
| $\delta = 4$ | Collagenase arm only | |
| | Surgery arm only | |
| | Both arms | |
| $\delta = 5$ | Collagenase arm only | |
| | Surgery arm only | |
| | Both arms | |
| $\delta = 6$ | Collagenase arm only | |
| | Surgery arm only | |
| | Both arms | |
| $\delta = 7$ | Collagenase arm only | |
| | Surgery arm only | |
| | Both arms | |
| $\delta = 8$ | Collagenase arm only | |

| | | |
|---|---|---|
| | Surgery arm only | |
| | Both arms | |
| $\delta = 9$ | Collagenase arm only | |
| | Surgery arm only | |
| | Both arms | |
| $\delta = 10$ | Collagenase arm only | |
| | Surgery arm only | |
| | Both arms | |
| $\delta = -1$ | Collagenase arm only | |
| | Surgery arm only | |
| | Both arms | |
| $\delta = -2$ | Collagenase arm only | |
| | Surgery arm only | |
| | Both arms | |
| $\delta = -3$ | Collagenase arm only | |
| | Surgery arm only | |
| | Both arms | |
| $\delta = -4$ | Collagenase arm only | |
| | Surgery arm only | |
| | Both arms | |
| $\delta = -5$ | Collagenase arm only | |
| | Surgery arm only | |
| | Both arms | |
| $\delta = -6$ | Collagenase arm only | |
| | Surgery arm only | |
| | Both arms | |
| $\delta = -7$ | Collagenase arm only | |
| | Surgery arm only | |
| | Both arms | |
| $\delta = -8$ | Collagenase arm only | |
| | Surgery arm only | |
| | Both arms | |
| $\delta = -9$ | Collagenase arm only | |
| | Surgery arm only | |
| | Both arms | |
| $\delta = -10$ | Collagenase arm only | |
| | Surgery arm only | |
| | Both arms | |

**Table xx:** Brief baseline characteristics of participants by treatment received and allocation

| **Characteristic** | **Collagenase (N = xx)** | | **Surgery (N = xx)** | | **Total (N = xx)** | |
|---|---|---|---|---|---|---|
| | Received collagenase | Received other | Received surgery | Received other | Received allocated | Received other |
| **Age** | | | | | | |
| N | | | | | | |
| Mean (SD) | | | | | | |
| Median (Q1, Q3) | | | | | | |
| Min, Max | | | | | | |
| **Gender** | | | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| Male | | | | | | |
| Female | | | | | | |
| **Study reference joint** | | | | | | |
| MCP | | | | | | |
| PIP | | | | | | |
| **Number of comorbidities** | | | | | | |
| N | | | | | | |
| Mean (SD) | | | | | | |
| Median (Q1, Q3) | | | | | | |
| Min, Max | | | | | | |
| **Number of medications** | | | | | | |
| N | | | | | | |
| Mean (SD) | | | | | | |
| Median (Q1, Q3) | | | | | | |
| Min, Max | | | | | | |
| **PEM score at baseline** | | | | | | |
| N | | | | | | |
| Mean (SD) | | | | | | |
| Median (Q1, Q3) | | | | | | |
| Min, Max | | | | | | |
| **PEM score at treatment** | | | | | | |
| N | | | | | | |
| Mean (SD) | | | | | | |
| Median (Q1, Q3) | | | | | | |
| Min, Max | | | | | | |

# Appendix F – Secondary outcome measures

**Table xx:** Descriptive summaries of participant reported outcomes

| | Collagenase (N = )<br>N, Mean (SD) | Surgery (N = )<br>N, Mean (SD) | Total (N = )<br>N, Mean (SD) |
|---|---|---|---|
| **PEM overall questionnaire** | | | |
| 3 Months | | | |
| 6 Months | | | |
| 12 Months | | | |
| 24 Months | | | |
| **PEM treatment questionnaire** | | | |
| 3 Months | | | |
| 6 Months | | | |
| 12 Months | | | |
| 24 Months | | | |

**Table xx:** Descriptive summaries of participant reported outcomes and estimated differences between groups at each time point from fitted analysis models

| | Trial Arm: N, Mean (SD) | | Estimated difference (95% CI) |
|---|---|---|---|
| | Collagenase (N = ) | Surgery (N = ) | |
| **URAM** | | | |
| Baseline | | | |
| 3 Months | | | |
| 6 Months | | | |
| 12 Months | | | |
| 24 Months | | | |
| **MHQ** | | | |
| Baseline | | | |
| 12 Months | | | |
| 24 Months | | | |
| **SANE** | | | |
| Baseline | | | |
| 2 Weeks | | | |
| 6 Weeks | | | |
| 3 Months | | | |
| 6 Months | | | |
| 12 Months | | | |
| 24 Months | | | |

**Table xx:** Overall hand assessment descriptive summaries

| | Collagenase (N = ) | Surgery (N = ) | Total (N = ) |
|---|---|---|---|
| **3 Months** | | | |
| Cured | | | |
| Much better | | | |
| A little better | | | |
| The same | | | |
| A little worse | | | |
| Much worse | | | |
| Terrible | | | |
| Missing | | | |
| **6 Months** | | | |
| Cured | | | |
| Much better | | | |
| A little better | | | |
| The same | | | |
| A little worse | | | |
| Much worse | | | |
| Terrible | | | |
| Missing | | | |
| **12 Months** | | | |

| | | | |
|---|---|---|---|
| Cured | | | |
| Much better | | | |
| A little better | | | |
| The same | | | |
| A little worse | | | |
| Much worse | | | |
| Terrible | | | |
| Missing | | | |
| **24 Months** | | | |
| Cured | | | |
| Much better | | | |
| A little better | | | |
| The same | | | |
| A little worse | | | |
| Much worse | | | |
| Terrible | | | |
| Missing | | | |

**Table xx:** Recurrence at 6 and 12 months post treatment

| Analysis set | Time point | Collagenase ($n_1/m_1$) | Surgery ($n_2/m_2$) | OR (95% CI) | ARR (95% CI) |
|---|---|---|---|---|---|
| Goniometer only | Month 6 | | | | |
| | Month 12 | | | | |
| Supplemented | Month 6 | | | | |
| | Month 12 | | | | |
| Multiply imputed | Month 6 | - | - | | |
| | Month 12 | - | - | | |

**Table xx:** Descriptive summaries of active/passive RoM over time

| | Collagenase (N = ) | Surgery (N = ) | Total (N = ) |
|---|---|---|---|
| **Baseline** | | | |
| N | | | |
| Mean (SD) | | | |
| Median (Q1, Q3) | | | |
| Min, Max | | | |
| **Pre-treatment** | | | |
| N | | | |
| Mean (SD) | | | |
| Median (Q1, Q3) | | | |
| Min, Max | | | |
| **Month 3** | | | |
| N | | | |
| Mean (SD) | | | |
| Median (Q1, Q3) | | | |
| Min, Max | | | |
| **Month 6** | | | |
| N | | | |
| Mean (SD) | | | |
| Median (Q1, Q3) | | | |
| Min, Max | | | |
| **Month 12** | | | |
| N | | | |
| Mean (SD) | | | |
| Median (Q1, Q3) | | | |
| Min, Max | | | |
| **Month 24** | | | |
| N | | | |
| Mean (SD) | | | |
| Median (Q1, Q3) | | | |
| Min, Max | | | |

**Table xx:** Estimated differences in RoM at each post-treatment time point (Goniometer measurements only)

| | **Estimated difference (95% CI)** |
|---|---|
| **Active RoM** | |
| 3 Months | |
| 6 Months | |
| 12 Months | |
| 24 Months | |
| **Passive RoM** | |
| 3 Months | |
| 6 Months | |
| 12 Months | |
| 24 Months | |

**Table xx:** Estimated differences in RoM at each post-treatment time point (Goniometer and photographic measurements)

| | **Estimated difference (95% CI)** |
|---|---|
| **Active RoM** | |
| 3 Months | |
| 6 Months | |
| 12 Months | |
| 24 Months | |
| **Passive RoM** | |
| 3 Months | |
| 6 Months | |
| 12 Months | |
| 24 Months | |

**Table xx:** Estimated differences in RoM at each post-treatment time point (multiply imputed data)

| | **Estimated difference (95% CI)** |
|---|---|
| **Active RoM** | |
| 3 Months | |
| 6 Months | |
| 12 Months | |
| 24 Months | |
| **Passive RoM** | |
| 3 Months | |
| 6 Months | |
| 12 Months | |
| 24 Months | |

**Table xx:** Treatment complications by randomised group (worst single treatment complication experienced)

| **Level** | **Collagenase (N = )** | **Surgery (N = )** | **Total (N = )** |
|---|---|---|---|
| None (0) | | | |
| Technical – not complication (1) | | | |
| Very minor (2) | | | |
| Mild (3) | | | |
| Moderate (4) | | | |
| Severe (5) | | | |
| Devastating (6) | | | |
| Death (7) | | | |

**Table xx:** Estimated absolute differences in proportions experiencing complication of level j or higher, for j = 1,…, 7 from fitted proportional odds model

| **Covariate values (X)** | **Level** | **Absolute difference (95% CI) (Pr(Y ⩾ j \| collagenase, X) - Pr(Y ⩾ j \| surgery, X))** |
|---|---|---|
| Joint = MCP<br>Digit = Ring<br>PEM = mean | j = 1 (Technical) | |
| Joint = MCP<br>Digit = Little<br>PEM = mean | j = 1 (Technical) | |
| Joint = PIP<br>Digit = Ring<br>PEM = mean | j = 1 (Technical) | |
| Joint = PIP<br>Digit = Little<br>PEM = mean | j = 1 (Technical) | |
| Joint = MCP<br>Digit = Ring<br>PEM = mean | j = 2 (Very minor) | |
| Joint = MCP<br>Digit = Little<br>PEM = mean | j = 2 (Very minor) | |
| Joint = PIP<br>Digit = Ring<br>PEM = mean | j = 2 (Very minor) | |
| Joint = PIP<br>Digit = Little<br>PEM = mean | j = 2 (Very minor) | |
| | | |
| | | |
| | | |

**Table xx:** Odds ratios for allocation for each level j = 1,…,7 from partial proportional odds model

| **Level** | **Odds ratio for allocation (95% CI)** |
|---|---|
| j = 1 (Technical) | |
| j = 2 (Very minor) | |
| j = 3 (Mild) | |
| j = 4 (Moderate) | |
| j = 5 (Severe) | |
| j = 6 (Devastating) | |
| j = 7 (Death) | |

# Appendix F – Tables for Reporting Adverse Events

**Table X:** Serious adverse events

| ID (SAE ref#) | Site[1] | Description | Outcome | Relatedness | Expectedness | Allocation | Treatment received at time of event? | Time since treatment (days) |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |

**Table X:** Non-serious adverse events

| ID (AE ref#) | Site[1] | Description | Outcome | Relatedness | Expectedness | Allocation | Treatment received at time of event? | Time since treatment (days) |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |