The University Of Sheffield. Clinical Trials Research Unit.

MissionEB Statistical Analysis Plan V2 30Apr2024

Study Title:	**Mesenchymal Intravenous Stromal cell InfuSIONs in children with recessive dystrophic Epidermolysis Bullosa (MissionEB)**
Funder (Reference Number):	**NHS England/National Institute for Health Research (NIHR127963)**
Sponsor (Reference Number):	**Great Ormond Street Hospital for Children NHS Foundation Trust (18CB01)**
ISRCTN Trial Registration Number:	**ISRCTN14409785**
EudraCT Number:	**2020-005049-18**
Research Ethics Reference Number:	**21/NE/0016**

[redacted]

List of abbreviations

AE(s)	Adverse event(s)
SAE(s)	Serious adverse event(s)
BCH	Birmingham Children's Hospital
BM-MSCs	Bone marrow-derived mesenchymal stromal cells
CC	Complete case
CHU	Child Health Utility
CI	Confidence interval
CONSORT	Consolidated Standards of Reporting Trials
CrI	Credible interval
CRF	Case Report Form
CRF	Clinical Research Facility
CTRU	Clinical Trials Research Unit
DF	Dose-finding
DMC	Data monitoring committee
DMEC	Data monitoring and ethics committee
EBDASI	Epidermolysis Bullosa Disease Activity and Scarring Index
ELISA	enzyme linked immunosorbent assay
FA	Full analysis population
GCP	Good Clinical Practice
GOSH	Great Ormond Street Hospital
GvHD	graft-versus-host disease
hbd	highest posterior density
HEAP	Health Economics Analysis Plan
IDD	Internal dose de-escalation
IIF	indirect immunofluorescence
IMP	Interventional Medical Product
IQR	Interquartile range
iscorEB	instrument for Scoring Clinical Outcomes of research for Epidermolysis Bullosa
max	Maximum
MCMC	Markov chain Monte Carlo
MCSE	Monte-Carlo Standard error
ML	Maximum likelihood
min	Minimum
mITT	Modified intention-to-treat
MSCs	Mesenchymal stromal cells
NHS	National Health Service
NCI CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
NIHR EME	National Institute for Health Research Efficacy and Mechanism Evaluation
OR	odds ratio
QP	Qualified Person
PP	Per-protocol
RD	Risk difference
RDEB	Recessive dystrophic epidermolysis bullosa
REML	Restricted maximum likelihood
RR	Risk ratio / relative risk
SAP	Statistical Analysis Plan

SD	Standard deviation
SE	standard error
SOP	Standard Operating Procedure
SUSAR	Suspected Unexpected Serious Adverse Reaction
TMG	Trial management group
TSC	Trial Steering Committee
UC-MSCs	Umbilical cord-derived mesenchymal stromal cells

Trial summary

Trial title	**M**esenchymal **I**ntravenous **S**tromal cell Infu**SION**s in children with recessive dystrophic **E**pidermolysis **B**ullosa (Mission-EB).
Trial design	Double-blind, randomised, placebo-controlled, crossover trial with an internal dose de-escalation safety study.
Investigational medicinal product and dosage	Third-party umbilical cord-derived mesenchymal stromal cells (UC-MSCs), 2-3 million cells/kg; adjusted to 1-1.5 million cells/kg, if necessary, based on observed toxicity data during the internal dose de-escalation phase. UC-MSCs will be administered intravenously through infusions.
Comparator product	Placebo - non-active UC-MSCs infusion to mask blinding.
Trial participants and setting	Eligible participants from the 2 national centres for treating children with recessive dystrophic epidermolysis bullosa (RDEB): Great Ormond Street Hospital (GOSH) and Birmingham Children's Hospital (BCH).
Sample size	Internal dose de-escalation study: 4 + 5 cohort of children with RDEB for the dose de-escalation study based on feasibility. With potential for an additional 4 + 5 cohort if one or more of the first 3 treated children experiences one or more suspected unexpected serious adverse event (SUSAR) within 48 hours of infusion or 2 or more of the first 6 treated children experience a SUSAR within 48 hours of infusion. See Figure 1 and section 3.4 for full dose de-escalation decision rules. Full trial: a feasible total of 36 children with RDEB, including children in the dose de-escalation study. If dose reduction is required, the children receiving the initial higher does will be re-invited to take part in the trial on a lower dose after washout period.
Treatment and Follow-up period	**Internal dose de-escalation trial:** • 2-3 million/kg UC-MSCs, adjusted to 1-1.5 million/kg, if necessary, according to observed toxicity data, • UC-MSCs (day 0) + UC-MSCs (day 14) or • Placebo (day 0) + placebo (day 14). **Main crossover trial:** • UC-MSCs (day 0) + UC-MSCs (day 14) followed by placebo (9 months) + placebo (9 months & 2 weeks) or • Placebo (day 0) + placebo (day 14) followed by UC-MSCs (9 months) + UC-MSCs (9 months & 2 weeks)

	Open-label study: Depending on the results of the main crossover trial, there will be a 12 month continued treatment non-randomised open-label follow-on study: • 2-3 million/kg UC-MSCs, adjusted to 1-1.5 million/kg, if necessary, according to toxicity data, • Day 0 and Day 0 + 2 weeks – UC-MSCs infusion, • 4 months and 4 months + 2 weeks - UC-MSCs infusion, • 8 months and 8 months + 2 weeks - UC-MSCs infusion, Analysis of the open-label study is not covered in this SAP.
Outcome (Internal dose de-escalation study)	Toxicity as defined by a patient experiencing one or more SUSAR as outlined in section 9 of the protocol within 48 hours of receiving an infusion.
Outcomes (Main crossover trial)	Judgements on the efficacy of UC-MSCs will be based on totality of evidence from both primary and secondary clinical outcomes. **Primary** Change in disease severity as measured by Epidermolysis Bullosa Disease Activity and Scarring Index (EBDASI)[1] global score summed across all 5 domains at 3 months post-infusion of UC-MSCs (from day 0). **Secondary** Full details of secondary outcomes are in section 5.2.2 and include change in disease severity measured by instrument for Scoring Clinical Outcomes of Research for Epidermolysis Bullosa (iscorEB [2]); change in quality of life; health economic analysis (covered in a separate HEAP); AEs and SAEs and safety bloods
Analysis	**Internal dose de-escalation study:** After each stage, the data monitoring and ethics committee (DMEC) will be provided with summary data for the number experiencing a SUSAR with assessment of severity and relatedness. All SAEs will also be reported. No other outcome data (e.g., clinical efficacy) will be shared with the DMEC at this stage. **Main crossover trial**: Continuous outcomes will be analysed using a mixed effects linear regression model with treatment, period, and baseline (for each period) (if necessary) as fixed effects and participant as a random effect. Treatment effect will be estimated using restricted maximum likelihood (REML) with Satterthwaite degrees of freedom.

	Equivalent Bayesian mixed-effects model will be fitted for EBDASI and iscorEB using weakly-informative priors on model parameters to estimate the probabilities of the treatment effect showing any positive improvement on selected outcomes. Binary outcomes and change in wound appearance (ordinal) will be analysed using the mixed effects logistic regression and mixed effects ordinal logistic regression models, respectively with period and treatment as fixed effects and participant as a random effect.

Table of Contents

1 Background and research rationale 12

2 Scope of the statistical analysis plan 12

3 Outline of the study design features 13

3.1 Rationale for the crossover trial with an internal pilot 13

3.2 Washout period and rationale 14

3.3 Starting dose of UC-MSCs and rationale 14

3.4 Internal dose de-escalation phase and adaptive features 15

3.5 2x2 Crossover trial 16

3.6 Follow-on open-label and uncontrolled study 17

4 Study aims and objectives. 17

4.1 Primary objectives 18

4.2 Secondary objectives relating to the main study. 18

5 Outcome measures 18

5.1 Internal phase I dose de-escalation trial 18

5.2 Main crossover trial 18

5.2.1 Primary outcome 19

5.2.2 Secondary outcomes 19

6 Sample size justification 20

6.1 Sample size for internal dose de-escalation study. 20

6.2 Sample size for main crossover trial 20

6.3 Sample size for open-label follow-on study 20

7 Randomisation, sequence concealment, and blinding 20

7.1 Randomisation process and concealment 20

7.2 Randomisation method 21

7.3 Blinding 21

7.4 Trial oversight and measures to minimise operational bias. 22

7.4.1 The Data Monitoring and Ethics Committee (DMEC) 22

7.4.2 Trial Steering Committee (TSC) 23

7.4.3 Trial Management Group (TMG) 23

8 Interim analysis, decision rules, and decision-making process 24

9 Data Sources 25

10 Study population 25
10.1 Inclusion criteria 25
10.2 Exclusion criteria 25
11 Analysis population for the dose de-escalation phase 26
12 Analysis populations for the crossover trial 27
12.1 Modified intention to treat (mITT) population 27
12.2 Full analysis (FA) population 28
12.3 Complete case (CC) population 28
12.4 Per protocol (PP) population 29
12.5 Safety population 29
13 Data reporting during the trial 29
13.1 General considerations 29
13.2 Internal dose de-escalation phase 30
13.2.1 General considerations 30
13.2.2 Unblinded report 30
13.2.3 Blinded report 30
13.2.4 Funder and sponsor reports 30
14 Outline of dose de-escalation analysis 30
14.1 General statistical considerations 31
14.2 Participant flowchart 31
14.3 Characterisation of participants in the IDD phase 32
14.4 Uptake of infusions 33
14.5 Analysis of the primary outcome 33
15 Outline of main crossover trial analysis 36
15.1 General statistical considerations 36
15.2 Criteria for claiming evidence and decision-making process 36
15.3 Estimands aspects 36
15.4 Considerations for a crossover design 37
15.5 The use of period baseline measures 38
15.6 Participant flowchart and Breeches/non-compliances 38
15.6.1 Participant flowchart 38
15.6.2 Breaches/non-compliances 39
15.7 Demographics and baseline characteristics of participants 39
15.8 Uptake of infusions 42

15.9 Overview of variable types and analytical models 43

15.9.1 Continuous variables 43
15.9.2 Ordinal variables 44
15.9.3 Binary variables 44

15.10 Analysis of change in disease severity (EBDASI and iscorEB) at 3 months 45

15.10.1 General considerations 45
15.10.2 Descriptive statistics and data visualisation 45
15.10.3 Analysis of change in disease severity at 3 months 48
15.10.4 Sensitivity analysis adjusting for period baseline 51
15.10.5 Comparing the primary and baseline adjusted models. 51
15.10.6 Bayesian model 52
15.10.7 Subgroup analysis by participant age 55
15.10.8 Sensitivity analysis using alternative populations 55
15.10.9 Sensitivity to timing of bloods 55
15.10.10 Additional considerations 55

15.11 Analysis of changes in disease severity (EBDASI and IscorEB scores) at 6 months 56

15.11.1 General Considerations 56
15.11.2 Summary statistics and charts 56
15.11.3 Analysis model and sensitivity 56

15.12 Analysis of changes in disease severity using EBDASI domain scores at 3 and 6 months 56

15.12.1 Clinical justification 56
15.12.2 Summary statistics and charts 57
15.12.3 Statistical analysis model 57

15.13 Analysis of pain symptoms using the Wong Baker FACES Score and VAS 57

15.13.1 General considerations 57
15.13.2 Statistical analysis model 59

15.14 Analysis of general clinical appearance of wounds 61

15.14.1 General considerations 61
15.14.2 Data visualisation and descriptive summaries 61
15.14.3 Discordance in treatment response between period within treatment sequence 63
15.14.4 Analysis model for change in wounds appearance (ordinal outcome) 64
15.14.5 Analysis of any observed improvement in wound appearance (binary outcome) 65

15.15 Analysis of itch symptoms using the Itch man scale 67

15.15.1 General Considerations 67
15.15.2 Analysis model 67

15.16 Analysis of itch symptoms using the Leuven itch scale 68

15.16.1 General Considerations 68
15.16.2 Descriptive summary statistics and data visualisation 68
15.16.3 Analysis of change in six Leuven itch domains 68

15.17 Analysis of changes in analgesia and itch medication 69

15.17.1 General considerations 69

15.17.2 Descriptive summary statistics and data visualisation 69
15.17.3 Analysis of change in pain and itch medications (yes or no) 70
15.18 Analysis of the CHU-9D Health utility 70
15.18.1 General Considerations 70
15.18.2 Descriptive statistics and analysis 71
15.19 Analysis of safety outcomes 71
15.19.1 General Considerations 71
15.19.2 Descriptive statistics 71
15.19.3 Data visualisation 71
15.19.4 Statistical analysis of adverse events 72
16 Summary of statistical analysis approach for each outcome 75
17 Further details of statistical methods and calculations 75
17.1 Participant discontinuation 75
17.2 Missing data and Imputation 76
17.2.1 Item non-response 76
17.2.2 Missing outcome data and Multiple Imputation (MI) 76
17.3 Checking statistical model assumptions 77
17.3.1 Continuous outcomes 77
17.3.2 Ordinal outcomes treated as continuous 78
17.3.3 Binary outcomes 78
17.3.4 Ordinal outcomes 79
17.3.5 Adverse events (with time offset) 79
17.4 Bayesian model diagnostics and convergence 80
18 Data manipulations and definitions 80
18.1 General considerations 80
18.2 The EBDASI 81
18.3 iscorEB 82
18.3.1 iscorEB questions completed by the clinician 82
18.3.2 iscorEB questions completed by the patient or parent/guardian 84
18.4 Wong Baker FACES score 84
18.5 Visual Analogue Scale for pain 84
18.6 The itch man scale (ages 4 to 13) 85
18.7 Leuven itch score (ages 14 and over) 85
18.8 Clinical appearance 88
18.9 Dose and frequency of pain medication 88
18.10 Dose and frequency of itch medication 89
18.11 CHU-9D 89

18.12 Adverse events (AEs) 89
18.13 Derivation of follow up date and per-protocol treatment windows. 90
18.14 Calculating follow up period for safety analysis. 90
18.15 Summary of questionnaire instruments 91
19 Implementation of the SAP 92
20 References 93
21 Appendix 95

1 Background and research rationale

Recessive dystrophic epidermolysis bullosa (RDEB) is a rare and inherited skin blistering disease affecting children where blistering follows minor injury because of a missing protein in the skin and other organs. Wound healing is often slow, leading to chronic erosions, secondary infection and progressing to extensive, mutilating scars and contractures. The disease has a significant medical, physical, emotional, and socio-economic impact on the patients and their families. Currently, there is no effective treatment for RDEB and the management of children with the disease is mainly supportive, with daily foam absorbent dressing changes often taking 1-4 hours to perform and topical and systemic antimicrobials for skin infections. Patients also require regular and breakthrough analgesics, including opioids several times at high and unlicensed doses. Furthermore, their care requires multidisciplinary professionals and is costly.

Preliminary results from the EBSTEM trial [3] (an uncontrolled, open-label study) of 10 children with RDEB suggested that treatment with intravenous mesenchymal stromal cells (MSCs) are safe and indicated early evidence of disease amelioration by improving the appearance of the wounds, reducing pain and itch, and improving quality of life for the children and their families [3]. As such, a robust controlled trial was required to validate these preliminary findings and add to the evidence base on the treatment of RDEB using umbilical cord-derived mesenchymal stromal cells (UC-MSCs), which are potentially more effective than bone marrow-derived (BM) MSCs. Therefore, this MissionEB trial was undertaken to provide robust and further evidence on the safety and benefits of treating children with RDEB using third party intravenous UC-MSCs compared to placebo. A detailed background is found in the trial protocol.

2 Scope of the statistical analysis plan

This statistical analysis plan (SAP) is written guided by the protocol, Good Clinical Practice (GCP), the Medicines for Human Use (Clinical Trials) Regulations 2004 [4], the International Conference on Harmonisation Statistical principles for clinical trials [5], appropriate standard operating procedures (SOPs) from the University of Sheffield Clinical Trials Research Unit (CTRU), available guidance for SAPs in clinical trials [6], and aspects from relevant reporting guidelines [7][8][4][9].

This SAP will guide the Trial Statistician(s) during the statistical analysis to answer the safety and clinical efficacy research objectives relating to the internal dose de-escalation (IDD) phase (safety interim analyses) and the overall crossover trial. The following aspects are outside the scope of this SAP and will be addressed elsewhere:

- health economics analysis to address additional health impacts and costs associated with UC-MSCs treatment (addressed in a standalone health economics analysis plan, HEAP);
- analysis of an open-label and non-randomised study to assess whether clinical benefits observed during the crossover trial (if any) are maintained over a 12-month period (addressed on a follow-on SAP following results from the main study);
- qualitative analysis to identify factors that are important for treatment acceptability and explore the impact of the MSCs on the lives of participating children and their families;
- analysis of research bloods that will be stored for further analysis in separate research.

All analysis will be performed with internal quality control in line with Sheffield CTRU SOPs (*https://ctru.ipassportqms.com/*) using validated statistical software such as Stata, SAS, R, or WinBUGS.

3 Outline of the study design features

Mission-EB is a double-blinded, randomised (1:1 to sequence), placebo-controlled, 2x2 (treatment by period) crossover trial with an internal phase 1 dose de-escalation trial for safety gatekeeping (Section 3.3). The trial investigates the safety and benefits of third-party UC-MSCs in treating children with RDEB. Efficacy of UC-MSCs will be determined from the totality of evidence from all clinical (primary and secondary) outcomes (Section 5).

3.1 Rationale for the crossover trial with an internal pilot

There are several challenges that influenced the study design. First, as highlighted in Section 1, RDEB is a rare condition, so the available sample size is very limited. Based on feasibility, the study is expected to recruit a total of 36 children with RDEB at two centres (Great Ormond Street Hospital (GOSH) and Birmingham Children's Hospital (BCH)). Therefore, there was a need for an appropriate study design that will generate robust evidence with a restricted population, and which utilises the available sample size more efficiently. Second, previous related studies at the time of the design were either open-label, uncontrolled or non-randomised; as such, the quality of evidence from these studies was very limiting. As a result, there was a need for a randomised and double-blinded trial to reduce potential biases and generate high quality evidence. A crossover trial where patients will be randomised to receive either placebo followed by UC-MSCs treatment or UC-MSCs treatment followed by placebo was viewed to be suitable (Section 3.2) and efficient as it allows for efficacy assessment with a fraction of the sample size compared to a parallel-group trial, each participating child acting as their own control, and all participating children receiving both study treatments at the end of the trial. Finally, there was a need for some safety gatekeeping of the chosen starting dose of UC-MSCs to ensure that is has acceptable toxicity profile before being administered to more

participating children. This necessitated the need for an internal pilot phase in the form of a randomised 4 + 5 dose de-escalation which would also be used as the first period treatments if the crossover trial proceeded (Section 3.5). Blinding aspects are discussed in Section 7.3 and are essential to ensure conscious or unconscious bias that may be experienced by study personnel does not impact on the conduct of the trial and its conclusions.

3.2 Washout period and rationale

There are necessary conditions for a crossover trial to be appropriate and for its efficiency to be realised. The design is not well-suited for treatments meant to cure the disease or when death during the duration of the trial is a likely outcome. Furthermore, there should be a sufficient time gap at the end of the first period to allow participants treated in that period to revert to their baseline clinical state (before treatment) after the end of treatment. Some issues are discussed in Section 15.

Currently, RDEB is an incurable condition and MSCs treatment is hoped to improve clinical outcomes of patients and their management rather than curing the condition. Based on results from an uncontrolled and open-label EBSTEM trial [3], clinical benefits peaked around 3 months from baseline, with these effects lasting for 4 to 6 months in most participants. Although this was based on 10 participants, it is clinically plausible to believe that these effects could last between 6 and 9 months to a lesser extent in a few participants. Consequently, the trial allowed for a 9 months' time gap between starting one treatment and crossing over to the start of another treatment in the second period so that by the start of the second period treatment participants will have reverted to their baseline state.

3.3 Starting dose of UC-MSCs and rationale

The UC-MSCs treatment will be started at dose level 2-3 million cells/kg as agreed by senior haematology clinicians. This cell dose was chosen based on the safety and efficacy data from EBSTEM trial [3] and previous clinical trials with intravenous BM-MSCs, mostly for steroid-resistant, graft-versus-host disease (GvHD) (see protocol Section 2.1.1). UC-MSCs have since emerged as a source of MSCs with several advantages over BM-MSCs. In previous related studies of BM-MSCs and UC-MSCs infusion treatment for GvHD, the cell dose level ranged from 3.4×10^5 to 7.2×10^6 per kg [10]. The EBSTEM trial [3] in RDEB patients used cell dose of 1-3 million cells/kg that was used in a related previous study in the Netherlands (NL13729.000.07) and no safety concerns were reported. See Section 2.2.1 of the protocol for more details including the rationale for the dosing schedule.

To further safeguard the safety of trial participants, the starting cell dose of 2-3 million/kg could be de-escalated to 1-1.5 million/kg depending on the observed safety profile of UC-MSCs treatment (Section 3.4).

3.4 Internal dose de-escalation phase and adaptive features

For safety gatekeeping, the IDD phase is a randomised 4 + 5 rule-based design with the safe dose being used as the first period treatments of the 2x2 crossover trial. The IDD study, summarised in the flowchart in Figure 1, will include a minimum of two stages and a maximum of four stages (2 cohorts of 4 + 5 patients) with continuation (cohort expansion), dose de-escalation, or early safety stopping decisions between stages. These decisions will be based on observed toxicities within 48 hrs of receiving an infusion (at day 0 or day 14) defined in Section 5.2.1. The first stage (block A cohort) will randomise four participants using a 3:1 ratio (UC-MSCs: placebo) with the higher MSCs dose (2-3 million/kg). Note that this allocation should be viewed as the first period treatment allocation of the 2x2 crossover trial (assuming it is found to have an acceptable safety profile). Randomisation of the next cohort of the subsequent stage based on the higher (2-3 million/kg) or lower (1-1.15 million/kg) UC-MSCs dose depends on toxicity outcomes observed from the first stage. For example, of the 3 participants randomised to UC-MSCs (2-3 million/kg), if 0 or 1 participant experienced toxicity, then following DMEC review and recommendation, the next cohort of 5 patients (block B) will be randomised using 3:2 ratio (UC-MSCs: placebo) on the same UC-MSCs dose level (2-3 million/kg). Otherwise, if 2 or 3 participants who were randomised to UC-MSCs (2-3 million/kg dose) experienced toxicities, then dose de-escalation will occur and the next cohort of 4 participants (block C) will be randomised using 3:1 ratio (UC-MSCs: placebo) on a lower UC-MSCs dose (1-1.15 million/kg). This process is repeated as shown in Figure 1.

If the safety / toxicity profile is found acceptable after block B or D, the study will proceed to the main crossover trial (Section 3.5) with an expectation to randomise 36 participants such that overall a 1:1 randomisation ratio is maintained (Section 7.1). Otherwise, the trial will be stopped early pending further view by the funder (NHS England/NIHR EME).

The strategy to randomise between UC-MSCs and placebo in the IDD stage allows the trial to maintain blinding and also allows for the identification of any issues with the placebo manufactured product.

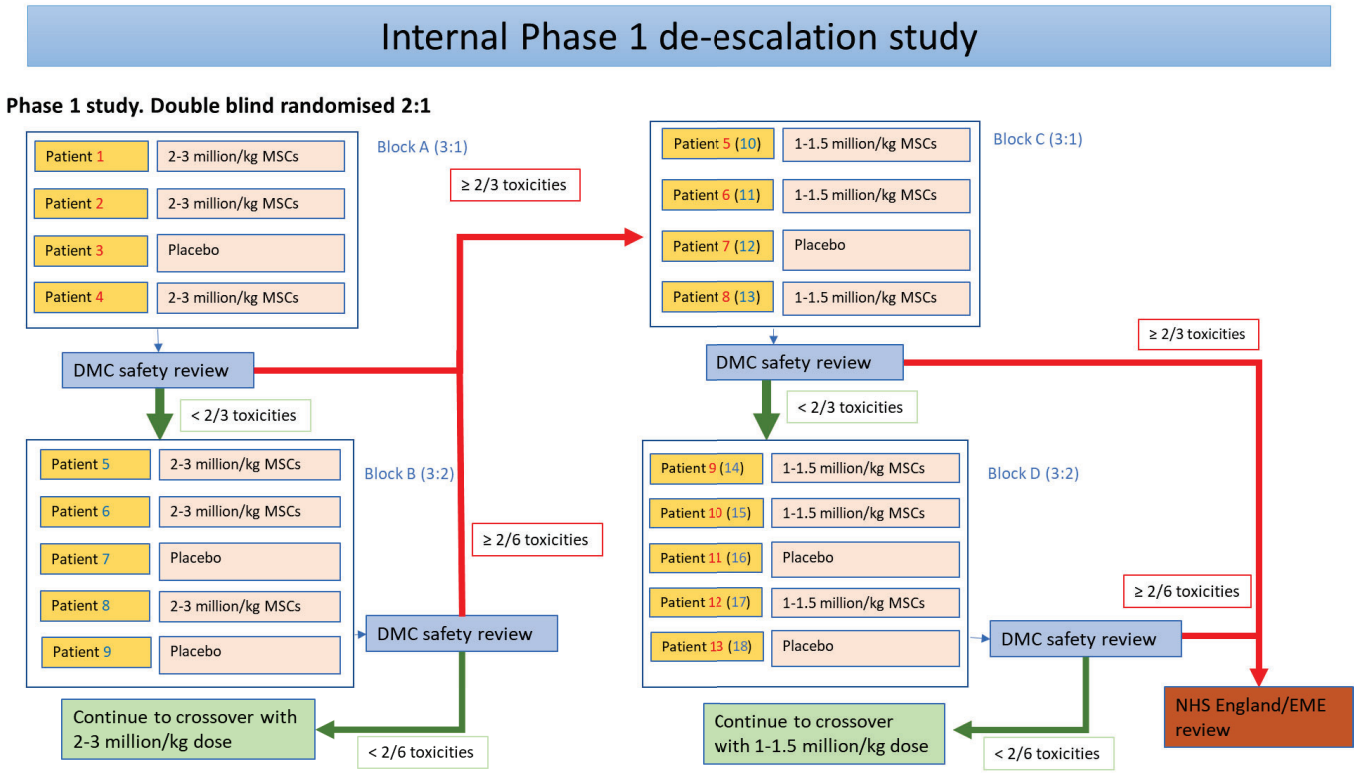

Figure 1. Flow diagram for the internal phase 1 dose de- escalation study.

3.5 2x2 Crossover trial

If the safety toxicity profile of the UC-MSCs dose is deemed acceptable following recommendation by the DMEC, the study will then progress to the full crossover trial using the recommended UC-MSCs dose. That is, patients will then be randomised to receive either a placebo followed by the UC-MSCs or the UC-MSCs followed by a placebo after a washout period. Participants who took part in the dose de-escalation phase in cohorts (blocks) that used the recommended dose will seamlessly transition to the crossover trial to receive their second period crossover treatments after the washout period of the first period treatments given during the IDD phase. Any participants that were randomised in cohorts using an unselected UC-MSCs dose will be expected to be re-randomised to restart their crossover trial after the dose de-escalation phase. Figure 2 summarises the crossover trial.

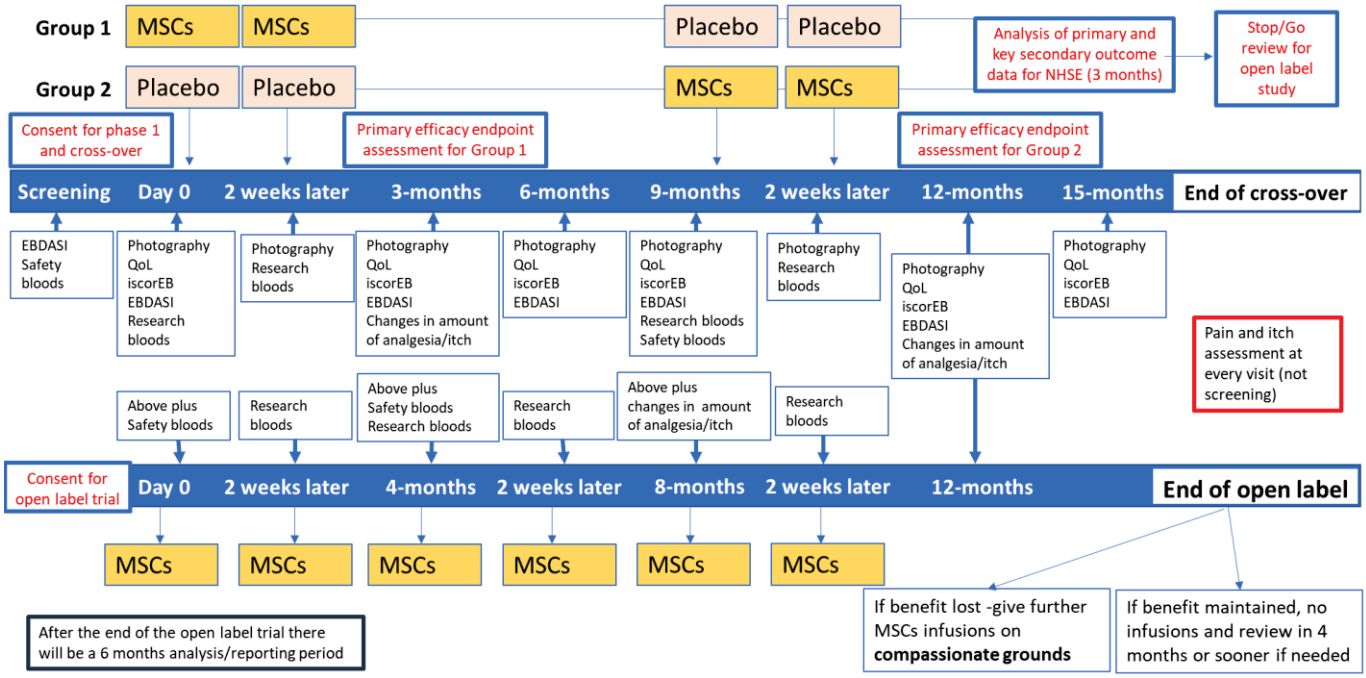

Figure 2. Flow diagram for the 2x2 crossover trial.

3.6 Follow-on open-label and uncontrolled study

If the treatment is found to be effective based on totality of evidence considering all outcomes in the crossover trial and there are no safety concerns (Section 13) an open-label study without a placebo administered but with all participants receiving UC-MSCs will commence one month after the end of the crossover trial. The objective is to assess whether the efficacy observed in the crossover phase (if any) is maintained over these 12 months.

All participants of the crossover trial, regardless of whether they have reached age 18 and moved to adult services, will be invited to take part in the open-label study and consent will be retaken. All participants will have been unblinded at the end of the crossover trial so they will know the period of the crossover trial in which they received the UC-MSCs treatment when they are invited to participant in the open-label study. Each participant will receive a total of six intravenous infusions at the dose established in the IDD study as summarised in Figure 1. Subsequently there will be one follow-up visit at 12 months. Full details of the analysis will be covered subsequently in a separate SAP.

4 Study aims and objectives.

The overall aim of the study is to assess if repeated infusions of UC-MSCs are safe and can benefit children with RDEB. Specific objectives relating to different components of the study are summarised in Sections 4.1 and 4.2.

4.1 Primary objectives

- Internal phase 1 dose de-escalation study: To assess the safety of third-party intravenous UC-MSCs in children with RDEB.
- Main study (crossover and open-label): To assess the efficacy of third-party intravenous UC-MSCs in improving disease severity in children with RDEB.

A separate SAP will cover the analysis of the open-label follow-on study depending on the results of the main crossover trial.

4.2 Secondary objectives relating to the main study.

1. To assess the safety of repeated UC-MSCs in children with RDEB,
2. To assess the efficacy of repeated UC-MSCs in improving quality of life and symptoms (e.g., pain and itch) in children with RDEB,
3. To undertake a health economic analysis to assess the costs and consequences of treatment with UC-MSCs versus usual care,
4. To explore patients' and parents' views in relation to treatment effectiveness and acceptability.

Analysis methods to address objectives 3 and 4 are not addressed in this SAP and will be addressed elsewhere.

5 Outcome measures

This section describes the outcomes for addressing objectives relating to different components of the study stated in Section 4, excluding those relating to health economics and qualitative aspects (stated in Section 2).

5.1 Internal phase I dose de-escalation trial

The primary outcome for the IDD phase is toxicity as defined by a patient experiencing at least one Suspected Unexpected Serious Adverse Reaction (SUSAR) within 48 hours of receiving their infusions at day 0 and day 14. The toxicity fraction relating to the number of patients who experienced at least one toxicity relative to the number of participants in the toxicity analysis population defined in Section 11 will be used to inform dose de-escalation, cohort expansion, early stopping for safety (Figure 1). Detailed definition of a SUSAR is provided in Section 9 of the protocol.

5.2 Main crossover trial

Judgements on the clinical efficacy and safety of UC-MSCs will be based on the totality of evidence from both primary and secondary clinical outcomes described in Sections 4.1 and 4.2. For all outcomes listed below, "day 0" baseline refers to the start of each period – that is, the start of the first period at

month 0 day 0 and, following the 3 month washout period after the first 6 months, the start of the second period at month 9 day 0 (Figure 2).

5.2.1 Primary outcome

Change in disease severity at 3 months post-infusion of UC-MSCs from day 0 (baseline) as measured by the Epidermolysis Bullosa Disease Activity and Scarring Index (EBDASI) [1] based on the total scores across all five domains.

5.2.2 Secondary outcomes

1) Change in disease severity at 6 months post infusion of UC-MSCs from day 0 (baseline) as measure by EBDASI [1] based on the total scores across all five domains. This is to assess the medium-term effect of treatment on disease severity as measured by EBDASI [1];
2) Changes in disease severity at 3 and 6 months post infusion of UC-MSCs from day 0 (baseline) as measured by the total score (combined from clinician and patient domains) of the Instrument for Scoring Clinical Outcomes of Research for EB (iscorEB [2]);
3) Changes in general clinical appearance of skin disease using clinical wound photography at 3- and 6-months post infusion of UC-MSCs from day 0 (baseline) as detailed in Section 15.14;
4) Changes in pain and itch as assessed by the Wong-Baker FACES Pain scale for children over 6 years old [11] and Leuven itch scale scores [12] at 3 and 6 months post infusion of UC-MSCs from day 0 (baseline);
5) Changes in amount of analgesia required for pain management at 3 months from day 0 (baseline) as assessed by whether the amount recorded within the last 48 hours has increased, decreased, or remained unchanged as detailed in Section 15.17;
6) Changes in the amount of itch medications at 3 months from day 0 (baseline) as assessed by whether the amount recorded within the last 48 hours has increased, decreased, or remained unchanged as detailed in Section 15.17;
7) Change in quality of life at 3 and 6 months post infusion of UC-MSCs from day 0 (baseline) using the validated Child Health Utility 9D (CHU-9D) scoring system [13] as detailed in Section 15.18;
8) Adverse events (AEs) and serious adverse events (SAEs) during the trial (see Section 9 of the protocol version 8 for definitions). Long-term AEs after the trial will also be collected as part of the open-label follow up study.

Secondary outcomes outside the scope of this SAP:

1. There is a separate HEAP for the health economic analysis,
2. A separate research application will cover the analysis of research bloods.

6 Sample size justification

6.1 Sample size for internal dose de-escalation study

The objective for the IDD phase is for safety gatekeeping rather than finding the optimal dose level of UC-MSCs treatment. The sample sizes for this safety gatekeeping are based on a 4+5 design, which is a variant of a 3+3 rule-based design with placebo controls to allow seamless transition into the main crossover trial. Placebo controls are to maintain blinding and to pick up any unexpected issues with the manufactured placebo that may occur. Furthermore, toxicity decisions are based on 3 participants in each cohort block that have received UC-MSCs infusions (Section 11).

6.2 Sample size for main crossover trial

As RDEB is a rare condition, the expected sample size of 36 is based on feasibility given the limited availability of the patients and not on formal power considerations. As such, the statistical analysis focuses on estimation rather than hypothesis testing. Table 1 gives the standardised widths for the precision of the trial (for a continuous outcome) as assessed by the half-width of a 95% confidence interval (CI) for different trial completion rates. For example, if 30 participants were administered all planned infusions and difference in primary outcome was available for both periods the confidence interval for the difference in the primary outcome between arms would be the estimated treatment effect plus or minus 0.53.

Table 1. Standardised widths for the precision of the trial.

Completed	Precision
36	0.49
30	0.53
25	0.59

6.3 Sample size for open-label follow-on study

The open-label follow-on study, when further treatment is deemed appropriate, will depend on available participants following the crossover.

7 Randomisation, sequence concealment, and blinding

7.1 Randomisation process and concealment

Site research staff at GOSH and BCH will randomise eligible participants with informed consent by entering their details using a computer via a validated web-based randomisation system hosted by the Sheffield Clinical Trials Research Unit (CTRU). This system has restricted access rights for allocation concealment such that the individuals or research staff who should not have access to the generated randomisation sequence are not able to do so. The trial statistician will generate the randomisation sequence using this web-based system by specifying the randomisation specifications but will not have

access to the generated randomisation sequence. A designated unblinded CTRU trial statistician independent of the day-to-day conduct of the MissionEB trial inspects, locks, and activates the generated sequence. The trial manager then activates recruitment so that site research staff can use the sequence for recruiting participants. This process is outlined in SOP08 on randomisation sequence generation.

7.2 Randomisation method

During the IDD phase, participants are randomised in four cohorts (see Figure 1) and the choice of the cohort block depends on toxicity decisions recommended by the DMEC after each cohort (Section 3.4). As shown in Figure 1, participants will be randomised in different ratios at different stages. The first 9 participants are divided into two cohorts all allocated using simple randomisation (blocked by cohort size); the first 4 are randomised (3: 1) to receive (UC-MSCs followed by placebo: placebo followed by UC-MSCs) and the next 5 patients are randomised (3: 2) to receive (UC-MSCs followed by placebo: placebo followed by UC-MSCs). This gives a (6: 3) overall allocation for these 9 participants to receive (UC-MSCs followed by placebo: placebo followed by UC-MSCs). It should be noted that the IDD phase uses only the 1st period treatments of the crossover sequence. Thus, these participants will only receive their 2nd period treatments of the crossover sequence after the washout period if no concerning toxicity issues are observed during the IDD phase using the 1st period treatments. It should be noted that dropouts prior to receiving their first infusion during the IDD phase will be replaced as outlined in a study related SOP08 on randomisation sequence generation.

If the safety profile during the IDD phase is deemed acceptable, the trial will proceed to the full crossover phase. At this stage, randomisation will proceed to achieve an overall 1:1 using blocked randomisation. Blocking was considered to ensure that the period sequences are balanced during the course of the trial to deal with the unpredictable nature of the COVID-19 pandemic. Only the trial and unblinded CTRU statisticians will know the block size during the trial. In overall, participants are randomised to achieve (1:1) allocation ratio to either receive UC-MSCs (1st period) followed by a placebo (2nd period) or placebo (1st period) followed by UC-MSCs (2nd period). All possible randomisation pathways and features across stages (from IDD phase to crossover phase) are detailed and document in a study document that will be publicly accessible during reporting.

7.3 Blinding

The knowledge by participants and research staff of what treatment has been administered can negatively influence trial conduct to produce biased results and misleading conclusions about the effects of study treatments. Blinding is therefore important to lessen this problem by limiting the knowledge of what treatments participants have received. The level of blinding required varies from trial to trial. MissionEB trial is a double-blinded study where all trial participants and the research team

involved in the day-to-day conduct of the trial as well as outcome assessors (e.g., wound photography) will be unaware of the treatment allocation. First, a placebo comparator using dimethyl sulfoxide (DMSO) – a preservative used in the MSCs suspension which has a characteristic odour and taste is used. Thus, the placebo comparator involves non-active infusions. Second, the treatment allocation is concealed as described in Section 7.1.

Intended unblinding will only occur after the crossover trial during the extended open-label study when required. To facilitate unblinding which may occur due to unforeseeable circumstances (e.g., safety) and manufacturing and packaging of the study treatments, a designated Sheffield CTRU statistician independent of the day-to-day conduct of the trial and the manufacturer (INmune Bio) will have secure access to the allocation sequence via the Sheffield CTRU validated web-based randomisation system. All unblinding incidences are recorded in the trial database. The designated unblinded CTRU statistician independent of the day-to-day trial conduct will create unblinded ongoing reports for monitoring by the data monitoring and ethics committee (DMEC) as described in Section 13. Site pharmacists, who are not involved in the direct conduct of the trial in any way are unblinded for qualified person (QP) inspection purposes of the Investigational Medicinal Product (IMP) during delivery. As the IMP/placebo will need to be thawed and prepared for administration, the suspension of cells will be apparent – thereby unblinding site research staff. To mitigate this, an independent research nurse from another team within the Clinical Research Facility (CRF) will prepare the IMP/placebo and administer infusions.

7.4 Trial oversight and measures to minimise operational bias.

There are three committees overseeing different aspects of the conduct of the trial. The interactions between these committees and other parties such as the funder, sponsor, and the Sheffield CTRU are detailed in Figure 3. The unblinded designated trial statistician independent of the running of the trial will be the only one with access to unblinded data for interim analysis.

7.4.1 The Data Monitoring and Ethics Committee (DMEC)

The DMEC oversees the welfare and safety of trial participants, reviews ongoing data, and makes recommendations to the trial steering committee (TSC) on dose de-escalation, cohort expansion, or early stopping for safety reasons (Section 3.4). The DMEC includes an independent chair, and its composition is detailed in the protocol and DMEC charter. The DMEC meets at regular intervals coinciding with interim safety reporting where relevant and as defined in the DMEC charter. In particular, the independent reviews of unblinded interim data to inform interim decisions after each cohort of patients in the IDD study (Figure 1) and during the crossover study (Figure 2) during closed meeting sessions. These unblinded reports, only shared with the DMEC, are prepared by the

designated Sheffield CTRU trial statistician(s) independent of the conduct of the trial as described in Section 7.3. In addition, the DMEC also reviews blinded interim reports during open meeting sessions, which are prepared by the data management team with oversight from the senior trial statistician.

7.4.2 Trial Steering Committee (TSC)

The TSC provides supervision of the protocol and statistical analysis plan, advises on and monitors progress of the study, reviews information from other sources and considers recommendations from the DMEC. The TSC includes an independent statistician, a paediatric dermatologist (Section 13.1 of protocol v8 and a patient representative; one of whom will be the independent chair. The committee meet at regular intervals, typically after the DMEC has met, and as defined in the TSC terms of reference. The TSC only reviews blinded ongoing/interim trial reports, which are prepared by the data management team with oversight from the senior trial statistician.

7.4.3 Trial Management Group (TMG)

The TMG is responsible for developing and implementing the trial protocol, day-to-day management and conduct of the trial, analysis, and reporting. The TMG consists of the chief investigator, principal investigators, and research team from Sheffield CTRU, GOSH and BCH. The CI will chair regular meetings to discuss the day-to-day running of the study, including any implementation issues. The TMG will review ongoing/interim blinded status reports throughout the trial but will not have access to any unblinded interim reports intended for the DMEC only. These TMG blinded interim reports are prepared by the data management team with oversight from the senior trial statistician who is otherwise not involved in the trial. See Section 13 for details of open (blinded) and closed (unblinded) reports.

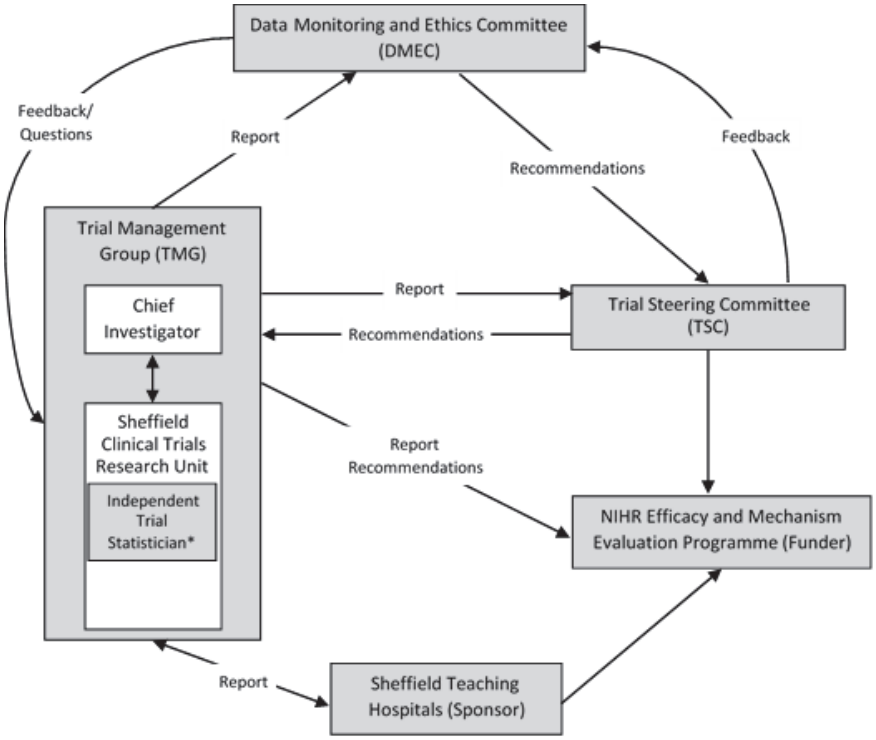

Figure 3. Interaction between Sheffield CTRU, committees, and other parties.

8 Interim analysis, decision rules, and decision-making process

As reflected in the protocol, the trial has an (embedded) IDD phase with between two and four stages. interim reporting after each stage (Figure 1), a fixed sample size for the crossover trial, an interim review of safety and efficacy at 3 months for the crossover trial (Section 13.2.4), and one formal statistical analysis at the end. After each stage (cohort block), randomisation will be paused and the DMEC provided interim data for a safety review as detailed in the DMEC charter and in line with agreed DMEC report template. If the number of toxicities from the initial dose exceed planned acceptable safety thresholds, new participants will be allocated using the same ratio and receive the lower UC-MSCs dose level. If the number of toxicities from the lower dose exceed agreed thresholds, the trial may be stopped. As highlighted in Section 7.4.1, the DMEC will make dose de-escalation, cohort expansion, or early trial stopping recommendations to the TSC. It will be the responsibility of the TSC to make decisions based on the recommendations from the DMEC on behalf of the sponsor and funder. Depending on the decision after the IDD phase, the funder and sponsor may review the dose de-escalation results (Figure 1). Figure 3 shows how different parties interact during decision-making process.

During the crossover phase, the DMEC will monitor allocation, retention, and safety by treatment arm with requests for more detailed monitoring information provided by another Sheffield CTRU statistician independent of this trial to preserve blinding. If there is evidence of harm due to the

intervention the trial may be stopped. At the end of the IDD phase, NHS England / NIHR EME will receive a report containing all safety reports submitted to the DMEC.

9 Data Sources

Data collection processes for this trial are detailed in Section 15 of the protocol. For concealment and masking purposes, data relating to randomisation (e.g., treatment allocation and order of treatment allocation in sequence) will be held within the Sheffield CTRU web-based randomisation system. The rest of the data (from CRFs and standardised questionnaires) are stored in a validated, web-based, and secure trial database management system hosted within the CTRU. Trial data will be extracted from source documents (including case report forms and participant questionnaires) and entered onto the CTRU's in house data management system (PROSPECT). Data sources are linked together by unique screening identifier for each trial participant. The CTRU data management team in the Sheffield CTRU will validate and query electronic data for inconsistencies throughout the course of the trial (as stipulated in SOP DM005). The trial statistician also will conduct any additional appropriate validation checks where appropriate before the data lock and sign off (as guided by DM005 and DM012). Appendix 2 contains a table from the trial protocol summarising the assessments made and data collected at each participant assessment.

10 Study population

This section describes the target study population to be enrolled in the trial and exclusion criteria in line with recommended items for inclusion in a SAP [6].

10.1 Inclusion criteria

1. Patients who have a diagnosis of RDEB characterised by partial or complete C7 deficiency including generalised severe and generalised intermediate subtypes;
2. Patients who are over 6 months and before their 16th birthday at time of consent;
3. Patients whose responsible parent/guardian has voluntarily signed and dated an informed consent form prior to the first study intervention. Whenever the minor child is able to give consent, the minor's assent will be obtained in addition to the signed consent of the minor's legal guardian.

10.2 Exclusion criteria

1. Patients with other subtypes of EB such as EB simplex, dominant dystrophic EB, EB junctional and Kindler EB;

2. Subjects who have received oral or topical corticosteroids for more than 7 consecutive days within 30 days of enrolment into this study, with the exception of the following steroids with non-systemic effects and intended to relieve oesophageal symptoms: oral viscous budesonide and inhaled fluticasone. Patients with a known allergy to any of the constituents of the investigational product.
3. Patients with a known allergy to any of the constituents of the investigational product.
4. Patients with signs of active infection that requires treatment with oral or intravenous antibiotics within 7 days of screening;
5. Patients with a medical history or evidence of active malignancy, including cutaneous squamous cell carcinoma;
6. Patients with BOTH a) positive C7 enzyme linked immunosorbent assay (ELISA), and b) a positive indirect immunofluorescence (IIF) with binding to the base of salt split skin at screening;
7. Patients who are pregnant or of child-bearing potential who are not abstinent or practicing an acceptable means of contraception, as determined by the Investigator, for the duration of the treatment phase;
8. Patients having received MSCs from any source in the last 9 months;
9. Simultaneous or previous participation in any interventional trial within 3 months before entering this trial but participation in simultaneous registry, and diagnostic trials during the trial is allowed.

11 Analysis population for the dose de-escalation phase

Toxicity analysis population should be defined to ensure that the toxicity profile of UC-MSCs infusions is not underestimated. Treatment allocation will be as received rather than randomised. The dose de-escalation, cohort expansion, early stopping decisions after each cohort will be based on 3 participants receiving UC-MSCs infusions (not placebo) as summarised in Figure 1. To be included in this toxicity analysis population in each cohort, participants should meet the following conditions:

a) Had received two intended infusions as per protocol within the allowed time windows (see Section 18.13) at day 0 and day 14;
b) Had received the first infusion (day 0), experienced toxicity and failed to receive the second UC-MSCs infusion for any reason, or;
c) Has received the first UC-MSCs infusion (day 0), not experienced toxicity and failed to receive the second UC-MSCs infusion for safety reasons to be reviewed by the blinded TMG members on case-by-case basis (e.g., COVID infection). These participants will be replaced as detailed in a trial specific SOP and will transition seamlessly into the crossover phase if the dose level

if not de-escalated during the IDD phase. Otherwise, they will be recycled back into the trial after the washout period.

Participants who do not receive the first infusion at day 0 for logistical reasons will not be given the infusion at day 14 and will not be included in the toxicity analysis population but will be invited to be re-randomised in the crossover study.

Participants will be excluded from the toxicity analysis population if they failed to receive the second infusion for logistical reasons. These will still be in the study and would transition into the crossover if the dose level is not de-escalated and study is continued beyond the IDD phase. These participants, participants who fail to receive any infusions and those in item c), but deemed not to meet criteria for inclusion in toxicity analysis population will be excluded in toxicity decisions and replaced during the conduct of the IDD phase as detailed in a related study specific SOP regardless of the treatment allocation. All participants excluded from the IDD phase toxicity decisions will be reported to the DMEC during their routine monitoring meetings.

12 Analysis populations for the crossover trial

For the main crossover trial, there are participant and period data aspects that should be considered when defining analysis sets as well as the objective of the analysis (e.g., clinical efficacy or safety) and the purpose of the results. For clinical efficacy, the primary analysis will be based on modified intention to treat as defined in Section 12.1. Additional sensitivity analyses will be performed based on complete case and per protocol analysis populations as defined in Sections 12.3 and 12.4. To address safety objective, safety analysis population is defined in Section 11.

12.1 Modified intention to treat (mITT) population

This will include eligible participants randomised with informed consent who have outcome data for both periods and period baseline data for at least one period. If the baseline measurement in one period is missing, we will deterministically impute it using the baseline measurement for the other period. This is a reasonable under the assumption that there is no carryover effect.

If a participant discontinues after 12 months, they will be retained in this population (if we have their outcome data for both periods). If a participant discontinues before 12 months, they will be excluded from this population (by definition because we will not have their outcome at 12 months) but be included the full analysis population and the associated multiple imputation (Sections 12.2 and 17.2.2) if we have data for at least one baseline.

Sequence allocation will be based on randomised treatment sequence regardless of what treatments were actually administered. For example, if a participant was randomised to placebo/MSCs but actually received MSC/placebo they would be included in the Placebo/MSCs sequence in the analysis. In the extremely unlikely event that this occurs, we will perform sensitivity analysis to this mITT

population using an additional as-treated population where such participants would be included in the treatment sequence received rather than randomised sequence. This will only be performed if there is at least one participant meeting the criteria and if this is the case the results will be included in the main results table (Section 10).

If the trial is stopped during the dose de-escalation phase due to a SUSAR resulting in death the scope of the SAP will be reduced and will only include analysis of the IDD phase (Section 14) and descriptive analysis of any outcome measures that have been collected.

Unrelated deaths will be treated in the analysis using a while alive policy because the trial results will be used by policy makers from a population level health care perspective. Specifically, if death occurs after 12 months, participants will be retained in this population (if we have their outcome data for both periods). And, if death occurs before 12 months, they will be excluded from this population (by definition because we will not have their outcome at 12 months) but be included in the full analysis population and the associated multiple imputation (Sections 12.2 and 17.2.2) if we have data for at least one baseline.

12.2 Full analysis (FA) population

Unlike the mITT, this will include eligible participants randomised with informed consent who have at least one baseline period even if they are missing one or both period outcomes. Sequence allocation will be based on randomised treatment sequence regardless of what treatments were administered. Participants who discontinued or died (unrelated to treatment) following baseline will be included in this population. Multiple imputation for missing data will be considered and if deemed appropriate it will be performed for this population and for outcomes relating to the EBDASI and iscorrEB only. The more discontinuations (or deaths) before 12 months the greater the chance of us observing inconsistency in the estimated treatment effects from the different analysis populations and the more likely that multiple imputation using the full analysis population will be deemed appropriate at the discretion of the senior statistician (Section 17.2.2). For example, 1 or 2 discontinuations or deaths will have negligible impact on the treatment effect we are estimating (unlikely to bias results even if we exclude them in the analysis) but, as deaths and discontinuations increase, the need for sensitivity analysis on their impact on the treatment effect increases.

Sensitivity analysis will also be performed using treatment allocation as treated if necessary.

12.3 Complete case (CC) population

This population will include all participants for whom we have baseline and outcome available in both periods for the outcome being analysed. For example, if we have data for baseline and 3 months post infusion for both periods but are missing data at 6 months post infusions for one or both periods, the

participant will still be included in the complete case population for the analysis at 3 months post infusion. As with the mITT population, we will include such participants irrespective of whether the administration of infusions and timing of data collection was per-protocol. Participants who discontinued or died at any point between baseline and 12 months will be excluded from this population (by definition). Participants who discontinued or died after 12 months will be considered eligible for this population in line with the while alive estimand strategy.

12.4 Per protocol (PP) population

This population will be the same as the complete case population, in the sense that all measurements must be available, but participants for whom treatment deviated from the protocol will be excluded. We will exclude participants from the PP population in any of the following circumstances:

1) participants who received less than 2 intended doses (0 or 1) within either period (MSC or placebo or both).
2) participants who received the second infusion (MSC or placebo) more than 5 days before or 14 days after the planned date (infusion 1 plus 14 days).
3) participants who received infusion 3 more than 0.5 months before or 8 weeks after the planned date (infusion 1 plus 9 months).
4) one or more follow up assessments occurred more than 14 days either side of the planned date (3 and 6 months after the first infusion in each period).

These planned dates are illustrated in the study procedure table in Appendix 2 and calculation of dates is outlined in section 18.13.

Participants will not be excluded from the PP population if the volume of infusion administered was outside the planned range of 2-3 million MSCs per kg.

12.5 Safety population

This will be based on treatment as received rather than treatment as randomised. In addition, for the participant's period data to be included, the participant should have received at least one infusion for that period.

13 Data reporting during the trial

13.1 General considerations

Ongoing/interim status reports will be produced by an unblinded CTRU statistician not otherwise connected with the trial as described in Section 7.4. The TSC and TMG will only receive blinded reports. DMEC meetings will include open and closed sessions. During open sessions, all attendees will discuss blinded reports. Blinded attendees will leave the meeting prior to the DMEC discussing the unblinded report in the final session of the meeting.

13.2 Internal dose de-escalation phase

13.2.1 General considerations

As reflected in the protocol and DMEC charter, the planned DMEC safety reviews occur after each block of 4 or 5 patients (Figure 1). In addition, all SUSARs will be reported to the DMEC and sponsor as soon as possible so safety recommendations can be made earlier if necessary. For example, if the first two patients receiving the active infusion experience a SUSAR it would not be appropriate to administer the same dose to the next patient to be randomised. The process for reporting is outlined in Section 9 of the protocol version 8.

In addition to the continuous monitoring and planned safety reviews, the DMEC will review the unblinded status report for safety during the closed sessions of scheduled (or unscheduled) meetings. All DMEC recommendations will be shared with the TSC along with appropriate supporting information while maintaining blinding of the TSC.

13.2.2 Unblinded report

This will include participant flow, baseline characteristics, primary toxicity outcomes, other safety data, recommendations and actions (Table 2, Table 3, and Table 4, respectively). Data will be provided separately for each block and overall. Sections and tables in the report will be clearly labelled to make it clear which outcomes have been reported previously, which are being reported for the first time and previous recommendations and decisions and the dates they were made.

13.2.3 Blinded report

Other than the actions recommended by the DMEC about how to proceed with the study (see Section 8), the unblinded IDD report will provide baseline information only (Table 2). Data will again be reported separately for each block.

13.2.4 Funder and sponsor reports

Decisions made during the IDD phase by the DMEC, TSC, and TMG after each cohort will be communicated to the funder (NIHR EME) and sponsor (GOSH for Children NHS Foundation Trust). In addition, the DMEC, TSC, and funder will review 3 months safety and efficacy outcome data for the crossover study once this data is available for both periods to inform decisions about continuity of the open-label study to minimise study delays.

14 Outline of dose de-escalation analysis

This section describes the key considerations, analysis methods, and reporting of the safety outcome (described in Section 5.1) of participants who took part in the IDD phase. This will also cover any

supplementary safety data presented to the DMEC to help them make informed dose de-escalation, expansion, or early stopping recommendations.

14.1 General statistical considerations

There are several considerations when reporting early phase trials with dose escalation or de-escalation components which are addressed in the Consolidated Standards of Reporting Trials (CONSORT) Dose-finding Extension (DEFINE) guidance [14] and extended guidelines for early trial statistical analysis plans [9]. The IDD phase will be reported in line with this CONSORT-DEFINE Extension focusing on what was planned and what happened during the trial. Specifically, starting dose and justification (Section 3.3); participant flow and characteristic of participants who took part in each cohort, starting dose, dosing options, cohort sizes, maximum sample size, and planned interim decisions (Section 3.4, Figure 1); interim decisions made; interim results on toxicity that supported interim decisions; interim decision-making process (Section 8) any deviations from planned decision rules.

The rest of this section provides illustrative tables based on a mocked-up example in which the dose was de-escalated after block 1 and full data for block 2 has recently been submitted.

14.2 Participant flowchart

The flow of participants in the IDD phase from screening to assessment of toxicity will be summarised in a flowchart similar to Figure 4.

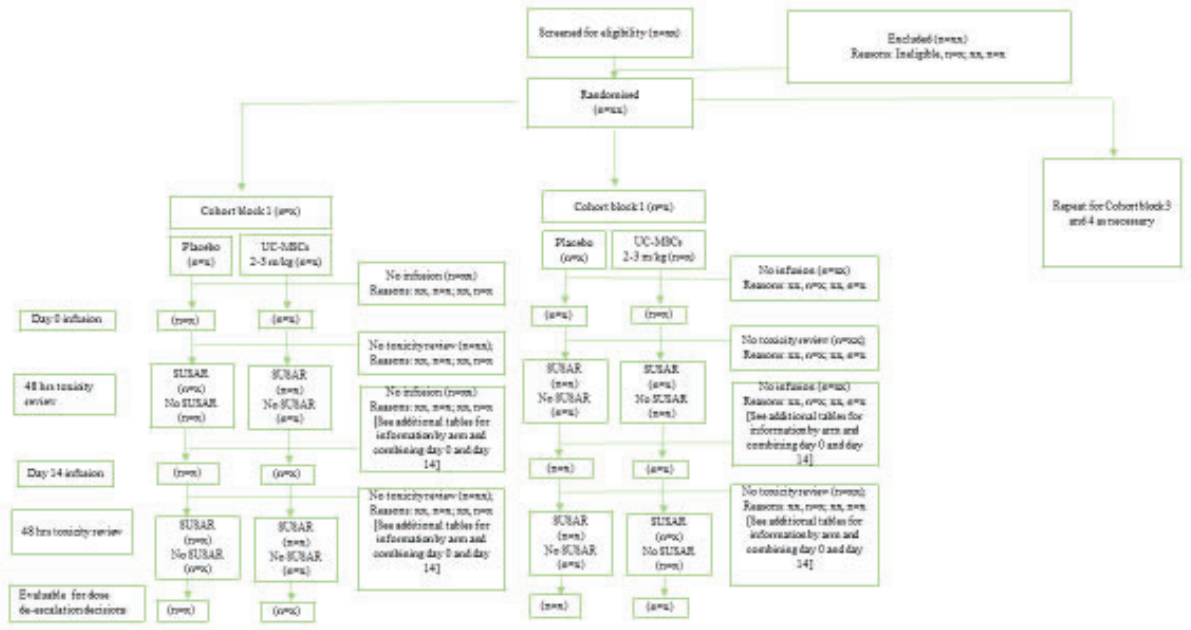

Figure 4. Participant flow chart for internal dose de-escalation study.

14.3 Characterisation of participants in the IDD phase

It is important for consumers of research findings to understand the nature of participants who took part in the IDD phase in order to put the results into context. To achieve this, key baseline characteristics of participants who took part in each cohort block will be listed as illustrated in Table 2.

Table 2. Baseline characteristics of participants in dose de-escalation study to date.

Block	Participant number (order	Allocation[1]	Site	Dose[2]	Age (years)	Sex (MF[3])	Ethnicity[4]	Type of RDEB[5]
1	1	T	GOSH	2-3	xx	xx	xx	xx
1	2	T	BCH	2-3	xx	xx	xx	xx
1	3	T	…	2-3	xx	xx	xx	xx
1	4	P		n/a	xx	xx	xx	xx
2	5	T		1-1.5	xx	xx	xx	xx
2	6	T		1-1.5	xx	xx	xx	xx
2	7	T		1-1.5	xx	xx	xx	xx
2	8	P		n/a	xx	xx	xx	xx

1 T = Treatment, P=Placebo; 2 million / Kg MSCs; 3 M= Male, F = Female; 4 English / Welsh / Scottish / Northern Irish / British, Irish, Gypsy or Irish Traveller, Any other White background, Indian, Pakistani, Bangladeshi, Chinese, Any other Asian background, White and Black Caribbean, White and Black African, White and Asian, Any other mixed / multiple ethnic background, African, Caribbean, Any other Black / African / Caribbean background, Arab, Any other; 5 Recessive Dystrophic EB (RDEB), EB simplex, Dominant DEB, Junctional EB, Kindler EB.

14.4 Uptake of infusions

The description of the uptake of treatment and whether the treatment was administered as per protocol give context to the safety and efficacy results. The reasons for failure to receive infusions will be presented in the IDD flowchart (Figure 1). In addition to this, the time between day 0 and day 14 infusions and whether infusions were done as per protocol will be listed as shown in Table 3.

Table 3. Uptake of Infusions.

Block	Participant	Allocation	Infusion 1 (day 0)	Infusion 2 (day 14 ±3 days)	Time between infusions (days)	As per protocol
1	1	T	Yes/no			Yes/no
1	2	T	…			…
1	3	T				
1	4	P				
2	5	T				
2	6	T				
2	7	T				
2	8	P				

14.5 Analysis of the primary outcome

The objective of this phase is to offer safety gatekeeping of the proposed dose based on the assessment of toxicity data that relates to all SUSARs due to study treatment as deemed by the study clinicians. Toxicity analysis population is defined in Section 11. The toxicity fraction (as defined in Section 5.1) is the one that influences the interim recommendations by the DMEC as described in Section 3.4 and Figure 1. Any toxicities observed in the placebo participants if any will also be reported side by side. Finally, the interim recommendations by the DMEC and final decision made by the TSC will be reported as shown in Table 4.

Details of any expected SAEs and AEs that were presented to the DMEC during the interim reporting period will be reported as supplementary information in a similar table to Table 4. These include, but are not limited to, immediate reactions such as severe allergic reactions, severe hypoxia, and severe shortness of breath, and/or chest pain. As reflected in Section 8, the DMEC will assess these data and recommend whether to continue with the proposed dose, halve the proposed dose and continue with the dose de-escalation study, proceed to the main crossover trial or stop the trial if the proposed dose

if deemed unsafe. Detailed decision rules are in Figure 1 of the protocol. AEs and SAEs by treatment group will be reported using descriptive statistics. Monitoring of safety data by the DMEC including toxicities will continue throughout the trial.

We will report the number of toxicities as a fraction of the number of active infusions in each block, and overall, as a cumulative fraction of all infusions at each dose as illustrated in Table 4.

Table 4. Observed toxicities and interim decisions made.

Block	Participant	Allocation	Dose (range)	Dose (actual)	Toxicity observed within 48hrs after day 0 infusion	Toxicity observed within 48hrs after day 14 infusion	Cumulative number of toxicities	Toxicity threshold	DMEC recommendation	TSC recommendation	TMG decision
1	1	T	2-3	2.47	N	N	0/0	2	Text[1]	Text	Text
1	2	T	2-3		Y	Y	1 / 2	2	…	…	
1	3	T	2-3		N	Y	2 /3	2			
1	4	P	n/a		n/a	n/a	n/a	2			
2	5	T	1-1.5	1.23	N	N	0/1				
2	6	T	1-1.5		N	N	0/2				
2	7	T	1-1.5		N	N	0/3				
2	8	P	n/a		n/a	n/a	n/a				

1 Cohort expansion, de-escalation, temporary pause or stop the trial early for safety (toxicity concerns)

15 Outline of main crossover trial analysis

This section outlines details of analysis methods of outcomes (described in Section 5.2.1) of participants who took part in the main crossover trial in order to address trial objectives described in Section 4. Complete details of data derivations are in Section 18. Further details of statistical methods are in Section 17.

15.1 General statistical considerations

The analysis methods and reporting should reflect the trial design addressing all outcomes and relevant information that facilitates the interpretation of results. With this in mind, we will follow the existing generic guidance on statistical analysis plans in clinical trials [6], but reflecting the crossover design used. Furthermore, the reporting of the crossover trial will adhere to the minimum essential standards in the CONSORT extensions for randomised crossover trials [7], for reporting of harms [15], and other aspects that we view as important to enhance the interpretation of the results.

15.2 Criteria for claiming evidence and decision-making process

MissionEB is not an adequately powered study for feasibility reasons explained above and in Section 3.1. There is no existing effective treatment and current care of children with RDEB is mainly supportive. RDEB is a naturally progressive condition affecting multiple aspects of children's health and wellbeing including their supporting families. As such, there is a challenging benefit-risk trade-off of potential treatments based on impact on multiple outcomes. In addition, currently, there is no clinical consensus to guide what is deemed to be clinically meaningful improvements in relevant outcomes. Given these challenges, the efficacy of UC-MSCs will be determined from the totality of evidence from all clinical (primary and secondary) outcomes. This will be guided by an independent review of all outcome data by the DMEC and the TSC who will consider whether the findings indicate evidence of improvement in outcomes. In addition to their recommendations, the funder will also review the results and make final decisions to approve continuation of the trial to an open-label non-randomised study.

15.3 Estimands aspects

The following aspects of the estimand should be well defined to inform statistical analyses and facilitate the interpretation of results (ICH E9(R1) [16]

1. treatment condition of interest (Section 1),
2. target study population (Section 10) and analysis sets defined in Sections 11 and 12,
3. study objectives (Section 4),
4. outcomes (Section 5),
5. dealing with intercurrent events such as missing data, not receiving allocated treatment, discontinuations, and deaths (Section12.1),

6. target population-level summary measure of the treatment effect.

On item 6, population-level summary measures of treatment effects can be classed as marginal or conditional effects depending on the statistical model used. The choice of the population-level summary measure depends on several considerations such as ease-of-interpretation, clinical meaningfulness relating to average patient benefit, and the purpose of the results (e.g., to inform policymaking, individual patients or both). These factors and the outcome used influence the choice of the statistical model and analysis approach of an outcome to obtain the desired treatment effect measure. MissionEB trial results are intended to inform policymaking focusing on whether the average effects of UC-MSCs infusion are clinically worthwhile if administered to children with RDEB similar to those enrolled in this trial. Therefore, the marginal treatment effect estimand is of primary interest.

For continuous outcomes (such as the change in EBDASI or iscorEB scores), the population-level summary measure of the treatment effect of interest is the mean difference in change between UC-MSCs and placebo groups. The mean difference summary measure is collapsible and as a result, the choice between a conditional and marginal statistical model is less important as both produce consistent results that can be interpreted in the same way.

However, for binary outcomes, such as whether the amount of analgesia has reduced (yes/no), the choice between a marginal and conditional statistical model can make a marked difference on treatment effects and interpretation depending on the chosen population-level summary measure and whether that summary measure is collapsible or non-collapsible. For example, the odds ratio (OR) is non-collapsible whereas the risk difference (RD, difference in proportions) and risk ratio/relative risk (RR) are both collapsible. With this in mind, the RD will be the primary population-level summary measure of the treatment effect and the RR and OR will also be presented side-by-by side to enhance interpretation of the effect of treatment.

15.4 Considerations for a crossover design

This section gives some general background to a crossover design. The appropriateness of the crossover design and rationale for the washout period is discussed in Sections 3.1 and 3.2, respectively.

There are several sources of variation that can influence patient outcomes in a two-treatment by two-period (AB/BA) crossover design these sources need separating from the effect of the treatment in the analysis:

1. a carryover effect occurs when the effects of the treatment allocated to participants in the first period are persistent and affect the outcome of the patients in the second period. That is when the second period of treatment for a given participant is affected by the treatment previously received in the first period by improving or worsening their outcomes. One way to

mitigate this problem is to have a sufficient washout period between treatments to allow the residual effect of the first period of treatment to be eliminated. Of note, the carryover effect in this trial, if it exists, is in one direction affecting the UC-MSCs to placebo sequence only.;

2. a period effect can occur if the severity of the condition changes over time independently of the intervention. For example, the effect of weather conditions on symptom severity;
3. a period by treatment interaction effect can occur if the effectiveness of the treatment varies over time;
4. a person effect will be present because we are taking repeated measurements on individuals at two different outcome timepoints.

As we are using an AB/BA crossover design, the sequence, carryover, and period by treatment effect (interaction) cannot be distinguished and separately estimated. MissionEB incorporates a sufficiently long washout period (Section 3.2) and under the assumption that there is no carryover, sequence, or interaction effect. As such, we do not include sequence as a covariate in statistical models when estimating the treatment effect (e.g., Section 15.10.3).

15.5 The use of period baseline measures

As highlighted in Section 3.5, baseline assessments will be done at the beginning of each period with the second done after the washout period. These period baseline assessments are for several reasons:

1) allow us to calculate the change in outcomes which are of importance to clinical interpretation;
2) offer an opportunity to assess the comparability of participants' characteristics at the start of each period which can be helpful in interpretation;
3) although adjusting for baseline in crossover trials is not as efficient as in parallel-group randomised trials, baseline measures can be useful to adjust for, especially when there are indications of marked differences in period baselines (e.g., due to deterioration of participants' health over time).

15.6 Participant flowchart and Breeches/non-compliances

15.6.1 Participant flowchart

A CONSORT style diagram for an AB/BA crossover trial will summarise the flow of participants from screening through to the end of the trial as illustrated in Figure 5. Specifically, after randomisation, the presentation will be done by treatment in each period as well as discontinuations with reasons prior to infusions (at day 0 and day 14) and follow-up (3 and 6 months for the first period and 12 and 15 months for the second period). Uptake of infusions at day 0 and 14 in each period will also be reflected. In addition, we will clearly indicate participants recruited:

a) to the IDD phase who seamlessly transition into the main trial at the original dose;

b) to the IDD phase who flowed into the main trial at a lower dose after a dose de-escalation, if appropriate, and;

c) directly into the main crossover trial.

Data availability will be reported at each follow-up for selected key endpoints relating to EBDASI and iscorEB.

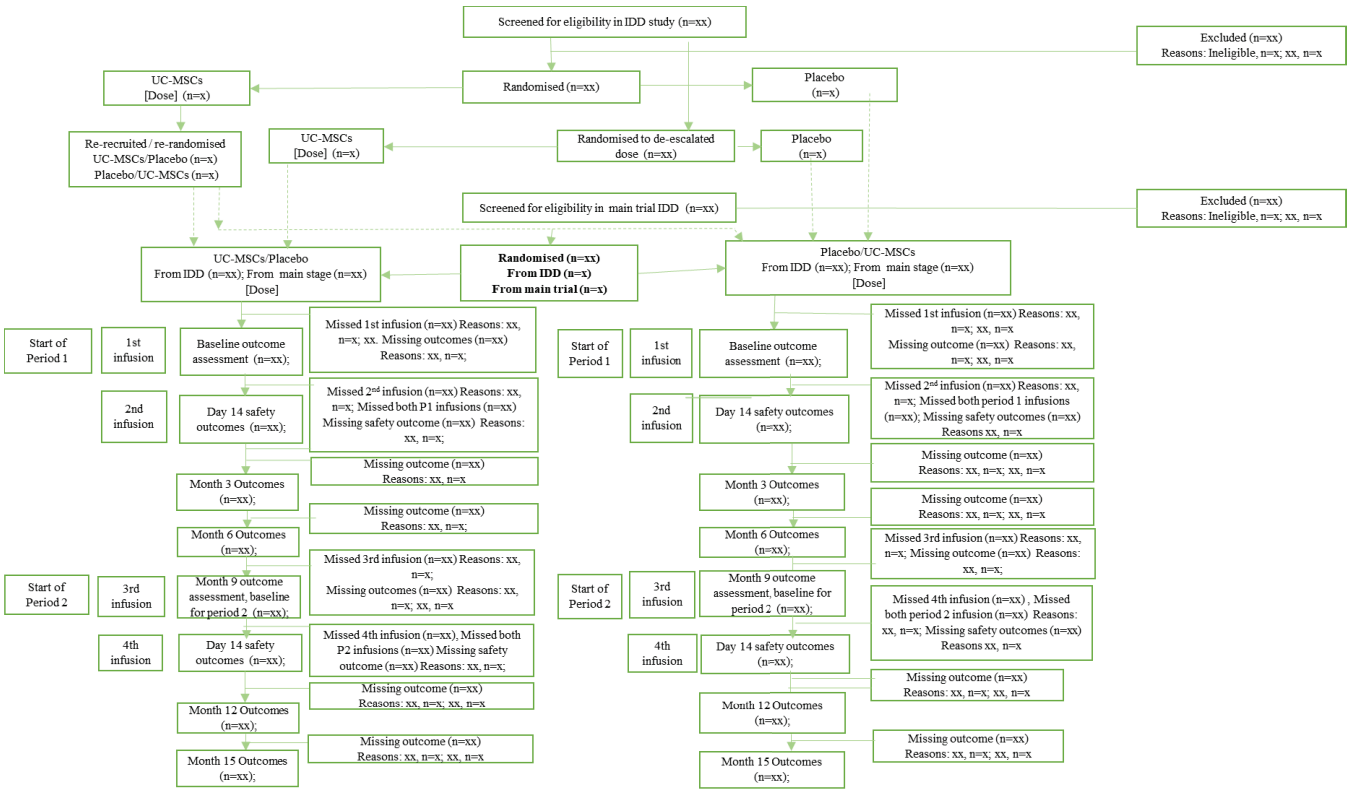

Figure 5. Participant flowcharts for IDD phase and main crossover trial.

15.6.2 Breaches/non-compliances

Breaches / non-compliances will be reported descriptively, overall and by grade (serious, major, minor). We will report total number; number of participants experiencing at least one; total number of each type and number of repeat occurrences per participant as applicable. Breaches / non-compliances are also recorded on a stand-alone excel spreadsheet will be reconciled with those entered into PROSPECT data management system.

15.7 Demographics and baseline characteristics of participants

In an AB/BA crossover design, each participant is expected to take part in each period once and receive only one of the treatments in each period such that after the end of the second period, they would have received both treatments. It is therefore important to report the overall characteristics of participants who took part in the trial and by randomised sequence. As described in Section 15.5, it is also important to describe baseline characteristics of participants at the start of each period for comparability. To address both issues, the characteristics of participants will be reported overall and

for each period by treatment group as illustrated in Table 5. Separate versions of the table illustrated in Table 5 will be produced for each analysis population (Section 12). Baseline characteristics for any recycled participants following a dose de-escalation will be based on data collected at the start of the main trial rather than data collected at the start of the IDD phase.

Summaries of continuous variables will show the number of participants, mean and standard deviation (SD) or median and interquartile range (IQR) as appropriate for the distributional form of the data, and minimum and maximum. Summaries of categorical variables will comprise the number of participants, and the number and percentage of participants in each category.

Table 5. Baseline characteristics of main crossover trial.

Demographics / Characteristics	Scoring	All participants	Sequence		Period 1		Period 2	
			MSCs/ Placebo	Placebo/ MSCs	MSCs	Placebo	MSCs	Placebo
		(n=xx)	(n=xx)	(n=xx)	(n=xx)	(n=xx)	(n=xx)	(n=xx)
Site	GOSH							
	BCH							
Ethnicity[5]								
	…							
Age (years) categorical	<10							
	≥10							
Age (years) continuous	Mean (SD)							
	Median (IQR)							
	Min, max							
Weight	Mean (SD)							
	Median (IQR)							
	Min, max							
BMI	Mean (SD)							
	Median (IQR)							
	Min, max							
Type of RDEB	Severe							
	Intermediate							
Other baseline characteristics of clinical interest[1]								
EBDASI overall score[2]	Mean (SD)							
	Median (IQR)							
	Min, max							
EBDASI damage score[2]	Mean (SD)							
	Median (IQR)							
	Min, max							
EBDASI activity score[2]	Mean (SD)							
	Median (IQR)							
	Min, max							
iscorEB overall score[2]	Mean (SD)							
	Median (IQR)							
	Min, max							
iscorEB clinician score	Mean (SD)							
	Median (IQR)							
	Min, max							

iscorEB patient score	Mean (SD)							
	Median (IQR)							
	Min, max							
Wong-Baker faces pain score (categorical)	0 (no hurt)	n=x (x%)						
	…[2,4,6,8]							
	10 (hurts worst)	n=x (x%)						
Wong-Baker faces score[2] (continuous)	Mean (SD)							
	Median (IQR)							
	Min, max							
Visual analogue scale: average pain[2]	Mean (SD)							
	Median (IQR)							
	Min, max							
Visual analogue scale: worst pain[2]	Mean (SD)							
	Median (IQR)							
	Min, max							
Itch man scale (categorical)	0 (no itch)	n=x (x%)						
	..[1, 2, 3]							
	4 (terribly)	n=x (x%)						
Itch man scale[2] (continuous)	Mean (SD)							
	Median (IQR)							
	Min, max							
Leuven Itch: itch frequency domain (categorical)	0 [never]	n=x (x%)						
	..[1,2,3]							
	4 [always]	n=x (x%)						
Leuven Itch: itch frequency score[2] (continuous)	Mean (SD)							
	Median (IQR)							
	Min, max							
Leuven Itch: itch duration domain (categorical)	0 [<30 mins]	n=x (x%)						
	…[1, 2]							
	3 [>2 hours]	n=x (x%)						
Leuven Itch: itch duration score[2] (continuous)	Mean (SD)							
	Median (IQR)							
	Min, max							
Leuven Itch: itch consequences score[2,3] (continuous)	Mean (SD)							
	Median (IQR)							
	Min, max							
Leuven Itch: itch severity score[2] (continuous only)	Mean (SD)							
	Median (IQR)							
	Min, max							
Leuven Itch: itch distress score[2] (continuous only)	Mean (SD)							
	Median (IQR)							
	Min, max							
Leuven Itch: itch surface area score[2] (continuous only)	Mean (SD)							
	Median (IQR)							
	Min, max							
CHU-9D[4]	Mean (SD)							
	Median (IQR)							
	Min, max							

SD = Standard deviation; IQR = Interquartile range; CHU-9D = Child Health Utility Score 9D.

1 Additional baseline characteristics might include: previous diagnosis through biopsy [Yes/No]; previous diagnosis through DNA analysis [Yes/No]; Type of C7 deficiency [partial/complete].

2 Lower scores indicate less severe symptoms.

3 Itch consequences score uses 12 categorical questions to calculate a continuous score. Counts for Individual questions will not be reported separately.

4 Lower scores indicate lower health utility

5 *English / Welsh / Scottish / Northern Irish / British, Irish, Gypsy or Irish Traveller, Any other White background, Indian, Pakistani, Bangladeshi, Chinese, Any other Asian background, White and Black Caribbean, White and Black African, White and Asian, Any other mixed / multiple ethnic background, African, Caribbean, Any other Black / African / Caribbean background, Arab, Any other.*

15.8 Uptake of infusions

Uptake of infusions will be reported in the CONSORT flow diagram (Figure 5) and the matrix table illustrated in Table 6. If the reasons for withdrawal or non-administration of infusions are too numerous to meaningfully fit into the CONSORT diagram, a summary will be included in the flowchart and separate tables providing full detail will be produced and referenced in the flowchart.

Table 6. Matrix table of infusions received.

	Infusion administered in Period 2											
	UC-MSCs/Placebo						**Placebo/UC-MSCs**					
		Both	Only Day 0	Only Day 14	Neither	Sum		Both	Only Day 0	Only Day 14	Neither	Sum
Infusion administered in Period 1	Both						Both					
	Day 0 only						Only Day 0					
	Day 14 only						Only Day 14					
	Neither infusion						Neither					
	Sum						Sum					

Table 7. Timing of infusions.

Participant	Sequence	Infusion		PP	Infusion		PP	Both PP
		1 (Day 0)	2 (Day 14)		3 (Days after Month 9 Day 0)	4 Days after dose 3		
1	UC-MSCs/Placebo	Y	15	Y	17	15	Y	Y
2		Y	43	N	100	14	N	N
3		-	-	n/a	n/a	14	Y	N
4		-						
5								
6		-						
..		. . . .						
36								

Table 8. Gap between Infusions by sequence and whether administered according to protocol.

	Number (%) administered as per-protocol		**Time lapse between infusions (days) – blue cells indicate protocol deviation.**				
	UC-MSCs/ Placebo	**Placebo/ UC-MSCs**		**USC-MCs/ Placebo**		**Placebo/ UC-MSCs**	
	n (%)	**n (%)**		**Period 1**	**Period 2**	**Period 1**	**Period 2**
All doses, all PP	x (xx%)	x (xx%)	Median (IQR)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
			Min, max	xx xx	xx xx	xx xx	xx xx
Received all doses, PP in period 1 only	x (xx%)	x (xx%)	Median (IQR)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
			Min, max	xx xx	xx xx	xx xx	xx xx
Received all doses, PP in period 2 only	x (xx%)	x (xx%)	Median (IQR)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
			Min, max	xx xx	xx xx	xx xx	xx xx
Received all doses, Neither period PP	x (xx%)	x (xx%)	Median (IQR)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)	xx (xx, xx)
			Min, max	xx xx	xx xx	xx xx	xx xx
Both doses P1 only, delivered PP	x (xx%)	x (xx%)	Median (IQR)	xx (xx, xx)		xx (xx, xx)	
			Min, max	xx xx		xx xx	
Both doses P1 only, NOT delivered PP	x (xx%)	x (xx%)	median (IQR)	xx (xx, xx)		xx (xx, xx)	
			Min, max	xx xx		xx xx	xx xx
Both doses P2 only, delivered PP	x (xx%)	x (xx%)	Median (IQR)		xx (xx, xx)		xx (xx, xx)
			Min, max		xx xx		xx xx
Both doses P2 only, NOT delivered	x (xx%)	x (xx%)	Median (IQR)		xx (xx, xx)		xx (xx, xx)
			Min, max		xx xx		xx xx
Both doses in neither period	x (xx%)	x (xx%)					
Total	(n=xx)	(n=xx)					

15.9 Overview of variable types and analytical models

This section summarises the three types of outcome variable in the trial and our analytical approach for each. The rest of Section 15 describes the analysis for each outcome in detail. Section 17.3 then describes how we will assess model validity for each type of outcome variable.

15.9.1 Continuous variables

Most of the outcomes are continuous or will be analysed as continuous for ease of interpretation. The continuous outcomes related to the trial are:

- EBDASI (Section 15.10 18.2);
- iscorEB (Section 15.1018.3);
- Visual Analogue pain scale (VAS, Section 15.13 18.4) and;
- three domains of the Leuven itch scale (Section 15.16).

Based on advice following consultation with the clinical team, the following ordinal variables will be analysed and interpreted as continuous variables:

- the Wong-Baker FACES pain scale (Section 15.13 18.4),

- the Itch Man scale (Section 15.15)
- three further domains of the Leuven Itch scale (Section 15.16).

Change in continuous variables compared to baseline will be analysed based on mITT population (Section 12.1) using a mixed linear effects model with treatment and period as fixed effects and participant as a random effect. Sensitivity analysis will include a model adjusting for period baseline and analysis of the complete case and a per-protocol populations (with and without baseline adjustment). Depending on the extent of missing data, we may also undertake analysis following multiple imputation based on FA population (Section 12.2). This same approach applies to 3 months and 6 months post infusion.

For EBDASI and iscorEB, we will also fit a Bayesian model with and without a baseline covariate at 3 months post-infusion. Full details of the continuous model are in the EBDASI section (Section 15.12) Details of model checking for continuous variables is in Section 17.3.1.

15.9.2 Ordinal variables

Change in wound appearance compared to period baseline is the only outcome that will be analysed as an ordinal variable. The primary analysis will be mITT using mixed effects ordered logistic regression with treatment and period as fixed effects and participant as a random effect (Section 15.14.4). Analysis of additional populations will follow the approach used for the continuous variables (Section 15.10.8). Details of model checking is in Section 17.3.4.

15.9.3 Binary variables

There are three binary outcome variables:

1) whether there was a decrease in analgesia compared to period baseline (yes/no) (Section 15.17);
2) whether there was a decrease in itch medication compared to period baseline (yes/no) (Section 15.17) and;
3) whether there was any improvement in the clinical appearance of wounds compared to period baseline (Section 15.14). As outlined above this is an ordinal outcome but will also be analysed as binary for ease of and supplementary interpretation.

The primary analysis of binary variables will be based on mITT using mixed effects logistic regression with treatment and period as fixed effects and participant as a random effect (Section 15.14.4). An equivalent mixed effects generalised linear model of the Poisson family with a log link function and robust standard errors (SEs) will be used to estimate treatment effect in terms of RR. Details of model checking is in Section 17.3.3.

15.10 Analysis of change in disease severity (EBDASI and iscorEB) at 3 months

This section covers two outcomes at 3 months post period baseline:

1. change in symptom severity using the overall score of the EBDASI (primary outcome),
2. change in symptom severity scores of the iscorEB.

These are considered together because they are continuous outcomes, so will be analysed in a similar fashion, and frequentist analysis will be complemented with a Bayesian analysis to aid interpretation of results. We illustrate the analysis using the change in overall EBDASI score but the same approach will be used for the change in iscorEB and the domains of the EBDASI.

15.10.1 General considerations

The EBDASI is a clinician reported outcome measure for assessing disease severity with an overall score that ranges between 0 and 506. Lower scores indicate less severe symptoms. It consists of two domains, activity and damage, with a maximum domain total score of 276 and 230, respectively. The iscorEB has one section completed by clinicians and another section completed by children or their parent/guardian. The two scores are summed to give an overall score that ranges from 0 to 258; maximum total scores for the clinical and patient domains are 138 and 120, respectively. Lower scores representing less severe symptoms.

Both measures also produce scores for various (sub)domains. See Section 15.12 for a consideration of the analysis of the domains. See Sections 18.2 and 18.3 respectively for further details of data collection and manipulation for the EBDASI and iscorEB.

15.10.2 Descriptive statistics and data visualisation

We will present the following summary statistics and charts:

- a table of summary statistics at period baseline and 3 months post infusion by period and sequence using mean, standard deviation (SD); median, quartiles, minimum, maximum, change in score at 3 months post-infusion compared to baseline and the between period difference for each sequence (Table 9).
- a line chart showing individual EBDASI scores at 0, 3, 9 and 12 months (Figure 6) and/or;
- a line plot showing change in EBDASI for each participant, split by sequence (Figure 7).

Table 9. Summary statistics for EBDASI by period and sequence.

	UC-MSCs/Placebo							Placebo/UC-MSCs							Overall Diff
	Period 1			Period 2			Diff*	Period 1			Period 2			Diff*	
	Day 0	+3mn	Change	Day 0	+3mn	Change		Day 0	+3mn	Change	Day 0	+3mn	Change		
Mean (SD)															
Median (IQR)															
Min, Max															

*Diff = Between period difference (UC-MSCs-Placebo)

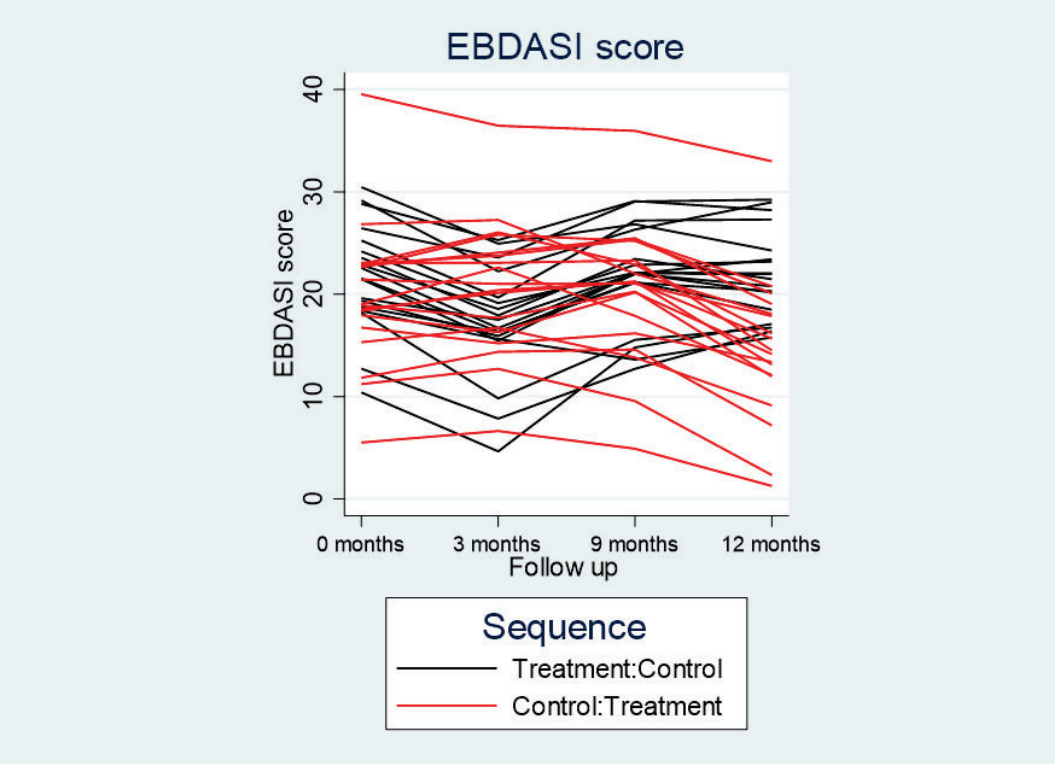

Figure 6. Line plot of EBDASI scores at baseline and 3 months follow-up for each participant.

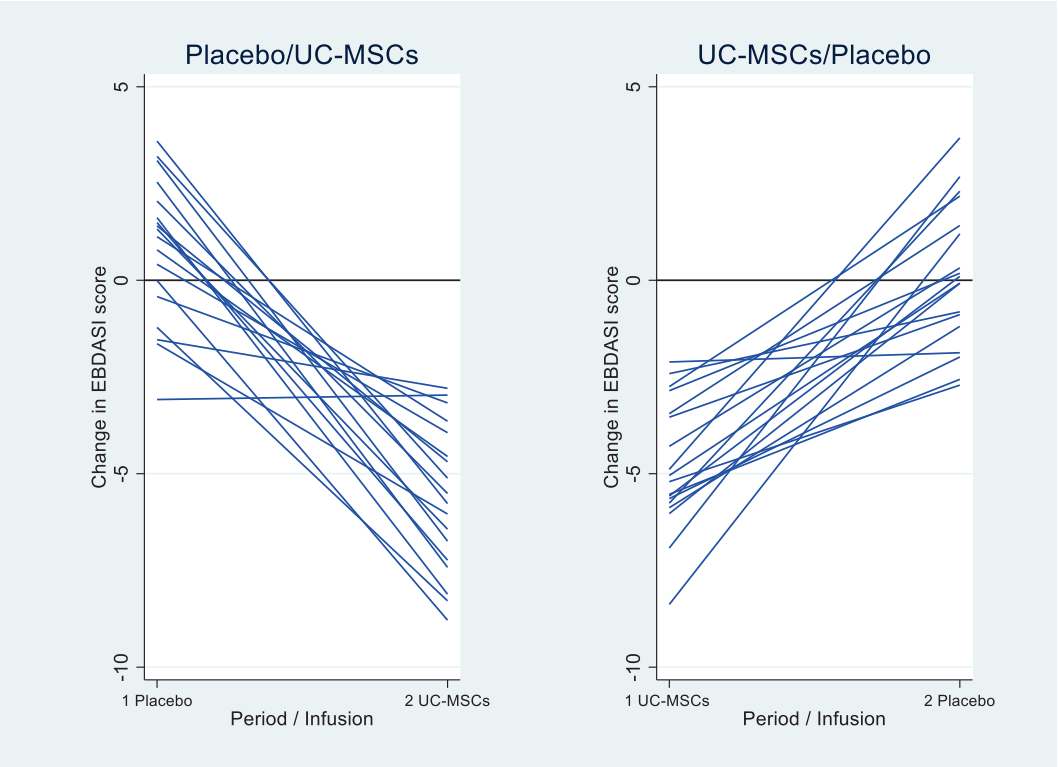

Figure 7. Line plot of each participant's EBDASI scores at baseline and 3 months follow-up by sequence.

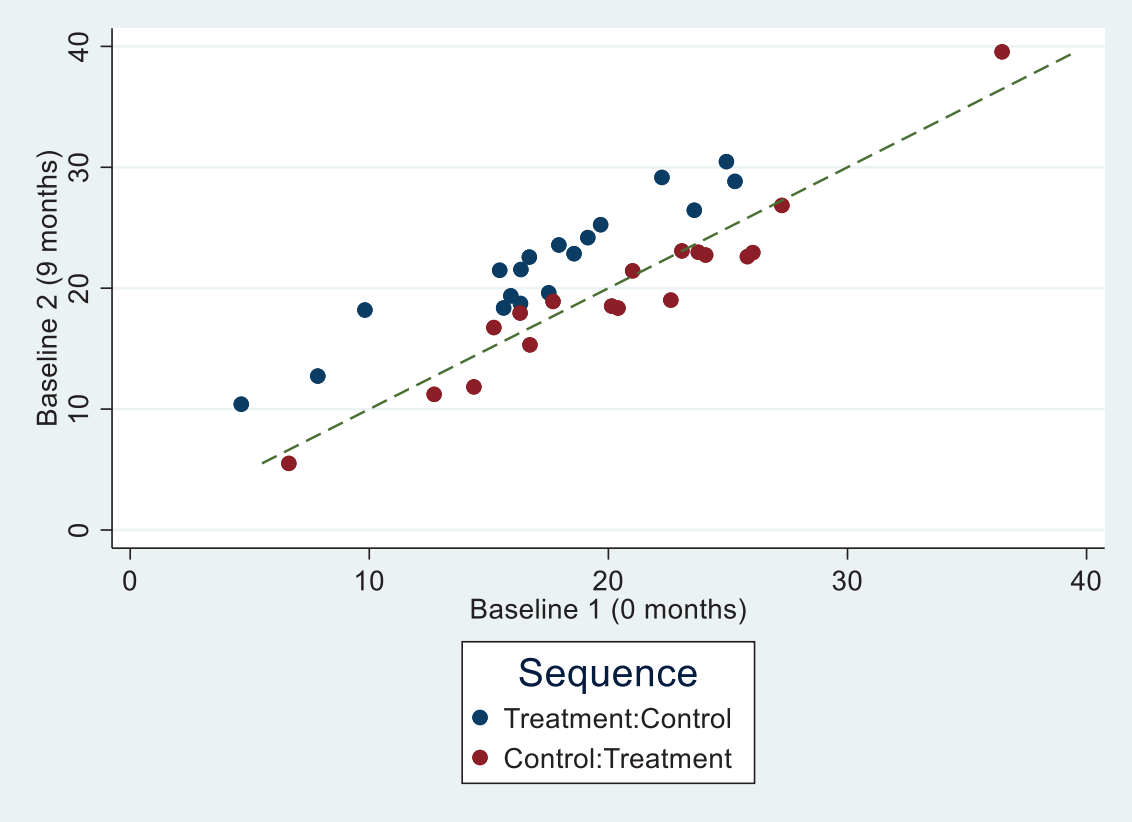

Figure 8. Scatter plot of period 1 baseline versus period 2 baseline.

15.10.3 Analysis of change in disease severity at 3 months

The primary analysis will use the mITT population (Section 12.1). The treatment effect will be estimated using a mixed effects linear regression model with treatment and period as fixed effects and a random subject/participant effect term. Restricted maximum likelihood (REML) with Satterthwaite degrees of freedom will be used to estimate the treatment effect and its SE. For an AB/BA crossover design, the mixed effects linear regression model can be mathematically represented by Equation 1:

$$y_{ijk} = b_0 + b_1 \text{Period} + \tau \text{Treat} + s_{ij} + \varepsilon_{ijk} \qquad \text{Equation 1}$$

Where:

a) y_{ijk} is the change in total scores for subject i {i=1, … ,n_j} of sequence j {j=1,2} in period k (k=1,2) 3 months from period baseline. The reason for this outcome is that results need to be presented in relation to change;

b) n is the number of people in each sequence; b_0 is the intercept; $b_1$is the period effect; τ is the marginal treatment effect of interest averaged across the two treatment sequences;

c) $s_{ij} \sim N(0, \sigma_a{}^2)$ is the subject/participant random effect for participant i of sequence j, which is assumed to be Normally distributed with mean 0 and variance $\sigma_a{}^2$, and;

d) $\varepsilon_{ijk} \sim N(0, {\sigma_e}^2)$ is the residual error for subject/participant i of sequence j in period k and the residual errors are assumed to be Normally distributed with mean 0 and variance ${\sigma_e}^2$.

This model can be fitted using the following Stata code when data are in long format and id is a participant indicator (id =1, 2, 3, …, n_1+n_2). There is only one random effect so an identity covariance structure will be used.

```
mixed outcome i.period i.trt || id: , dfmethod(sat) reml ///
        cformat(%3.2f) pformat(%4.3f) sformat(%5.3f)
```

The treatment effect will be presented as the adjusted mean difference in change with 95% CI. To understand the complete picture, the treatment effect in each sequence (UC-MSCs/placebo and placebo/UC-MSCs) that contributes to the overall treatment effect will be reported. Analysis will be repeated with an adjustment for period baseline (Section 15.10.4); for the PP populations (Section 12.4); for other analysis populations under certain conditions such as full analysis population (Section 12.2); using a Bayesian approach (Section 15.10.6). Results, excluding Bayesian outcomes, will be presented as shown in Table 10. The Bayesian results will be presented as shown in Table 11. In the event of substantial differences in period baselines as described in Section 15.10.2, the model described in Section 15.10.4 adjusted for baseline will be the primary analysis model.

Table 10. Difference between arms in the change in symptom severity measured using EBDASI and iscorEB.

Outcome	Analysis set	UC-MSCs/Placebo*		Placebo/UC-MSCs **		Paired adjusted mean difference in change (MSCs – Placebo) [95% CI] ‡	Paired adjusted mean difference in change (MSCs – Placebo) [95% CI] †
		n	LSM [95% CI]	n	LSM [95% CI]	No baseline adjustment	Baseline adjusted
Change in EBDASI at 3 months from period baseline	mITT (primary)	xx	xx(xx to xx)	xx	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
	CC	xx	xx(xx to xx)	xx	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
	PP	xx	xx(xx to xx)	xx	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
	FAS	xx	xx(xx to xx)	xx	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
Change in iscorEB at 3 months from period baseline	mITT (primary)	xx	xx(xx to xx)	xx	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
	CC	xx	xx(xx to xx)	xx	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
	PP	xx	xx(xx to xx)	xx	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
	FAS	xx	xx(xx to xx)	xx	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)

LSM = least squares mean

* Mean values less than 0 indicate an improvement in disease severity and favours UC-MSCs treatment; ** Mean values greater than 0 indicate an improvement in disease severity and favours UC-MSCs treatment; ‡ adjusted for a period fixed factor and subject/participant random effect; † adjusted for period baseline and period fixed factors, and subject/participant random effect from a model described in Section 15.10.3. Negative adjusted mean differences in change indicates beneficial effects of UC-MSCs in reducing disease severity compared to placebo.

15.10.4 Sensitivity analysis adjusting for period baseline

The primary analysis model in Section 15.10.3 assumes that a participant's baseline scores in period 2 is similar to their baseline score in period 1. Sensitivity to this assumption will be explored by using a similar model to the primary model with period baseline as an additional covariate. If there are indications of substantial differences in period baselines, for example due to deterioration in condition over time (Section 15.10.2), then this will be upgraded to be the primary analysis model.

The change in EBSASI and iscorEB scores at 3 months from period baseline will be modelled using a mixed effects linear regression model with period baseline, treatment and period as fixed factors and subject/participant random effect given by Equation 2:

$$Y_{ijk} = b_0 + b_1 baseline_{ijk} + b_2 Period + \tau Treat + s_{ij} + \varepsilon_{ijk} \qquad \text{Equation 2}$$

Where:

- $baseline_{ijk}$ is the period baseline for subject i {i=1, … ,n} of sequence j (i=1,2) in period k {k=1,2};
- b_1 is the baseline effect which is a regression coefficient of change in total scores given period baseline;
- b_2 is the period effect;
- other parameters are a specified from Equation 1.

The model can be fitted in Stata using the below code when data are in long format. The results will be presented in Table 10 alongside other results.

```
mixed outcome baseline i.period i.trt || id: , dfmethod(sat) reml ///
      cformat(%3.2f) pformat(%4.3f) sformat(%5.3f)
```

15.10.5 Comparing the primary and baseline adjusted models.

When comparing results from statistical models not adjusted for and unadjusted for period baselines, it is important to investigate the following:

- whether participant's scores have generally returned to a similar level observed at period 1 baseline by period 2 baseline. If this is not the case the model with baseline adjustment is preferable, and;
- whether the treatment effects between sequences are similar.

15.10.6 Bayesian model

As highlighted in Section 1, RDEB is currently an incurable and complex chronic condition. The clinical team believes that any improvement in outcomes would be helpful to children suffering from RDEB. As such, it is important to understand how likely the UC-MSCs treatment is to achieve any positive mean improvement in disease severity outcomes compared to placebo to provide more insight into the treatment effects to aid interpretation. As we are using the Bayesian model to aid interpretation of the frequentist results, the models should be as similar as possible. To achieve this objective, equivalent Bayesian mixed effects linear regression models described in Sections 15.10.3 and 15.10.4 will be used to analyse changes in EBDASI and iscorEB at 3 months post-infusion. The posterior distribution of the treatment effect will be estimated using Markov chain Monte Carlo (MCMC) sampling. The probability of the treatment achieving any positive mean improvement will be calculated from the resultant posterior distribution of the treatment effect. The posterior distribution of the treatment effect will be plotted for visual interpretation and the adjusted median difference in change between treatment groups (95% highest posterior density (hbd) credible interval (CrI)), adjusted mean in change with its SD and Monte Carlo SE will be reported as presented in Table 11. Five parallel MCMC chains will be simulated to check for MCMC convergence and to get more precise results pooled across the five chains using Gelman-Rubin's rules [17], [18]. A total of 50 000 MCMC iteration replicates will be used with a seed of 25048810 and 2500 number of iterations for the burn-in period. The number of iteration replicates could be increased depending on the results from Gelman-Rubin convergence diagnostic criteria (described in Section 17.4).

The prior distributions of all model parameters under considerations (described in Sections 15.10.3 and 15.10.4) need to be prespecified in the Bayesian model with rationale. There is no prior related outcome data from a controlled study to inform the choice of prior distribution and there is no basis to justify the use of informative prior. However, there is uncontrolled data from a small study of 8 participants with at least evaluable EBDASI or iscorEB data from GOSH based in BM-MSCs treatment (not UC-MSCs). A mean change (decrease) in scores of 13.0 and 5.9 from baseline average across variable follow-ups for each individual were observed for the EBDASI and iscorEB, respectively. The corresponding SDs were 7.4 and 22, respectively – however, the latter is unreliable due to an extreme outlier. With this in mind, for both outcomes, the treatment effect is assumed to be Normally distributed with mean 0 (under the null hypothesis) and a variance of 1000 reflecting huge uncertainty; an SD of 10 which is slightly larger than 7.4 (ignoring an unreliable iscorEB SD). For sensitivity analysis, the treatment effect will be assumed to be equally likely within plausible minimum and maximum values relating to change in EBDASI or iscorEB. That is, treatment effect is assumed to be uniformly

distributed with mean between values a and b. We chose values $[a, b]$ of [-20, 20] and [-13, 13] (for ebdasi and iscorrEB respectively) with conservative 7 points improvements more than observed in our small uncontrolled study. These two prior distributions of the treatment effect are non-informative in the sense that they do not favour the effect of treatment in one direction or the other and reflect huge uncertainty but within realistic thresholds.

As for other nuisance fixed effect parameters relating to the constant, period and period baseline (if specified in the model), their prior distribution is assumed to be the same and Normally distributed with mean 0 and variance 10^2 reflecting huge uncertainty. This is sensible under the assumptions of the crossover design given that there are no prior related crossover trials to inform their choice. For instance, we expect negligible period, sequence, and crossover effects. As for the nuisance random effect parameters relating to repeated measures and residual error, their prior distribution is assumed to the inverse gamma distribution. Using the notation of Equations 1 and 2, the prior distribution is specified below. The probabilities of the treatment showing any mean positive effect; i.e., the mean effect being less than 0 will be calculated by sampling from the posterior predictive distribution and presented as shown in Table 11. The code in the box below illustrates.

Prior distribution

$Y \sim N\ (b_0 + b_1 Period + \tau Treat,\ {\sigma_a}^2 + {\sigma_e}^2\ ); b_0, b_1, b_2, \sim N\ (0,\ 1000);$

${\sigma_a}^2, {\sigma_e}^2 \sim \Gamma^{-1}(0.01, 0.01)$, and $\tau \sim N\ (0,\ 1000)$ or $\tau \sim U\ (-a,\ a)$ in the sensitivity analysis.

Stata code without baseline in the model:

```
bayes, burn(2500) mcmcsize(50000) rseed(25048810)        ///
        nchains(5) hpd clevel(95)                        ///
        prior({outcome: period}, normal(0,1000))         ///
        prior({outcome: trt}, normal(0, 1000))           ///
        prior({e.outcome: sigma2}, igamma(0.01,0.01))    ///
        prior({U0:sigma2}, igamma(0.01,0.01)):           ///
 mixed outcome period trt || id:,
* compute the Gelman-Rubin summary measure for convergence checking
bayesstats grubin
* obtain posterior summary statistics in overall and by each chain
bayesstats summary
bayesstats summary, sepchains
* estimate posterior probability of any mean positive treatment effect
bayestest interval {outcome: trt}, upper(0)
```

Table 11. Bayesian estimates of adjusted differences and probability of achieving any improvement.

Outcome	Analysis set	No adjustment for baseline					Adjusting for baseline				
		Adjusted difference (UC-MSCs - Placebo)*				Probability of achieving any improvement	Adjusted difference (UC-MSCs - Placebo)*				Probability of achieving any improvement
		n	Median (95% CrI)	Mean (SD)	MCSE	Mean (MCSE) ‡	n	Median (95% CrI)	Mean (SD)	MCSE	Mean (MCSE) ‡
Change in EBDASI at 3 months from period baseline	mITT	xx	xx (xx, xx)	xx (xx)	xx	xx (xx)	xx	xx (xx, xx)	xx (xx)	xx	xx (xx)
	PP	xx	xx (xx, xx)	xx (xx)	xx	xx (xx)	xx	xx (xx, xx)	xx (xx)	xx	xx (xx)
	CC	xx	xx (xx, xx)	xx (xx)	xx	xx (xx)	xx	xx (xx, xx)	xx (xx)	xx	xx (xx)
	FA	xx	xx (xx, xx)	xx (xx)	xx	xx (xx)	xx	xx (xx, xx)	xx (xx)	xx	xx (xx)
Change in iscorEB at 3 months from period baseline	mITT	xx	xx (xx, xx)	xx (xx)	xx	xx (xx)	xx	xx (xx, xx)	xx (xx)	xx	xx (xx)
	PP	xx	xx (xx, xx)	xx (xx)	xx	xx (xx)	xx	xx (xx, xx)	xx (xx)	xx	xx (xx)
	CC	xx	xx (xx, xx)	xx (xx)	xx	xx (xx)	xx	xx (xx, xx)	xx (xx)	xx	xx (xx)
	FA	xx	xx (xx, xx)	xx (xx)	xx	xx (xx)	xx	xx (xx, xx)	xx (xx)	xx	xx (xx)

* Means/Medians less than 0 represent a decrease in disease severity symptoms and favour the treatment; MCSE (Monte-Carlo Standard Error); CrI (Credible interval); ‡ Prior distribution of treatment effect is $N\ (0,\ 1000)$. The table will be repeated using a uniform prior treatment distribution $U\ (-a,\ a)$.

15.10.7 Subgroup analysis by participant age

There is clinical interest in whether the effect of treatment on disease severity is consistent or homogeneous for participants aged less than ten compared to those aged 10 or over. We will explore this for the change in EBDASI and iscorEB scores at 3 and 6 months only by refitting the primary model as detailed in Section 15.10.3, but with a treatment by age group interaction term. Age group specific treatment effects and 95% CIs will be presented alongside the overall treatment effect (e.g., in a forest plot). If there are too few participants in either age group to fit a reliable statistical model, we will not undertake this analysis but only present the treatment effect in age group with most participants. This will be decided at the discretion of the lead trial statistician undertaking analysis during unblinded statistical analysis.

15.10.8 Sensitivity analysis using alternative populations

We will explore the sensitivity of estimates to missing data using FA population using multiple imputation as appropriate (Sections 12.2 and 17.2.2). If there are participants who deviated from the per-protocol treatment regime, we will explore the sensitivity of estimates to non-compliance (Section 15.6.2) by repeating the analysis for the PP population (see Section 12.4).

15.10.9 Sensitivity to timing of bloods

One subdomain of the clinical domain of the iscorrEB is the laboratory abnormalities score (0-15) which is scored using bloods. Due to difficulties in taking bloods, including the discomfort involved for participants, there is clinical discretion around whether bloods are taken every time the iscorrEB is scored. Bloods taken at alternative times can be used instead. For example, bloods taken at screening can be used as long as they are within 6 months of the scoring and bloods taken as part of routine care can also be used.

IscorrEB is a validated measure and contemporary bloods were most likely used in the validation. The use of bloods taken prior to iscorEB scoring might therefore affect its validity and make it less sensitive to improvements. To explore this possibility, we will calculate the time between the bloods being taken and the bloods being used to score the iscorrEB and plot this time against the score for the laboratory abnormalities domain.

At the discretion of the senior statistician, if the plot suggests there is a relationship, we will do sensitivity analysis by repeating the primary analysis for the subset of participants where contemporary bloods were used.

15.10.10 Additional considerations

Section B.1 on the iscorrEB assigns a score between 0 and 2 depending on the extent to which a participant can open their mouth. A clinical decision has been made that all children younger than 3 will be scored zero for this aspect of iscorrEB because it is not possible to consistently

make this assessment for children of this age because they will not have all developed incisors to the same extent. We will report the number of children that were below 3 (and hence automatically received a score 0) alongside summary measures of iscorrEB so this can be taken account of when the results are being interpreted.

15.11 Analysis of changes in disease severity (EBDASI and IscorEB scores) at 6 months

This section deals with change in EBDASI and iscorEB total scores at 6 months from period baseline to assess the medium-term effect of the treatment. Prior evidence from an uncontrolled small trial EBSTEM suggested that the effect of UC-MCs could last for 6 months or beyond as reflected in the protocol. This analysis aims to provide supporting evidence on whether this is the case.

15.11.1 General Considerations

Data collection and manipulations will be undertaken in the same manner as the change in score at 3 months and is summarised in Sections 18.2 and 18.3, respectively.

15.11.2 Summary statistics and charts

Descriptive statistics and tables will be the same as those presented for EBDASI at 3 months post baseline as illustrated in Section 15.10.2.

15.11.3 Analysis model and sensitivity

We will fit the same frequentist models as fitted for the change in score at 3 months post period baseline (Sections 15.10.3 and 15.10.4). To explore sensitivity to missing data we will use the same approach as outlined in Section 15.10.8 for the EBDASI at 3 months post period baseline.

15.12 Analysis of changes in disease severity using EBDASI domain scores at 3 and 6 months

Following advice from the clinical team, this section explores the effect of the treatment on change in total score of the two domains of the EBDASI at 3 and 6 months post period baseline:

1) change in symptom severity using the activity domain total score,
2) change in symptom severity using the damage domain total score.

Data collection and manipulations will be undertaken in the same manner as the change in score at 3 months and is summarised in Section 18.2.

15.12.1 Clinical justification

The clinical team believes that it is possible that the treatment may primarily affect disease activity with limited impact on disease damage. As such, the treatment may show more improvements in activity rather than damage domain.

15.12.2 Summary statistics and charts

Descriptive statistics and tables will be the same as those presented for change in EBDASI total scores at 3 months after baseline (Section 15.10.2).

15.12.3 Statistical analysis model

We will fit the same frequentist models as fitted for the change in score at 3 months post baseline, both with and without adjustment for baseline (Sections 15.10.3 and 15.10.4, respectively).

15.13 Analysis of pain symptoms using the Wong Baker FACES Score and VAS

This section covers the analysis of change in pain symptoms measured by the Wong Baker FACES scale and VAS at 3 and 6 months post period baseline.

The Wong Baker FACES scale is a child self-reported outcome collected for children aged seven and over. We are not collecting a child self-reported pain outcome for children aged six and under but the Visual Analogue scale for pain is collected for all ages (Section 18.5).

We illustrate the analysis of the Wong-Baker scale at 3 months, but the same approach will be used for analysis at 6 months.

15.13.1 General considerations

The Wong Baker FACES scale can be viewed as an ordinal outcome (with an underlying continuum). Children are shown a series of six faces and asked which one "best depicts the pain they are experiencing". Each ordinal response category has an associated numerical score 0 (no hurt), 2 (hurts little bit), 4 (hurts little more), 6 (hurts even more), 8 (hurts whole lot) and 10 (hurts worst). The clinical team has confirmed it is clinically meaningful to use the change in score associated with each category as a continuous outcome. The change in Wong Baker FACES scores will range from -10 to +10 with a negative score indicating an improvement in symptoms. Further details about data collection and derivation of the outcome are provided in Section 18.4.

The visual analogue scale (VAS) for pain is a continuous score between 0 and 10 reported by the child or their parents/guardians at 3 and 6 months post baseline. Thus, a change in VAS scores ranges from -10 to +10 with a negative score indicating improvement in pain symptoms.

These secondary outcomes at 3 and 6 months relates to the child's:

1) average pain in the last week,
2) worst pain in the last week.

For details of data sources and derivations see Section 18.5.

We will summarise the change in Wong Baker FACES and VAS scores as described in Section 15.10.2. Data will also be presented as shown in Figure 9, Figure 10 and Table 12

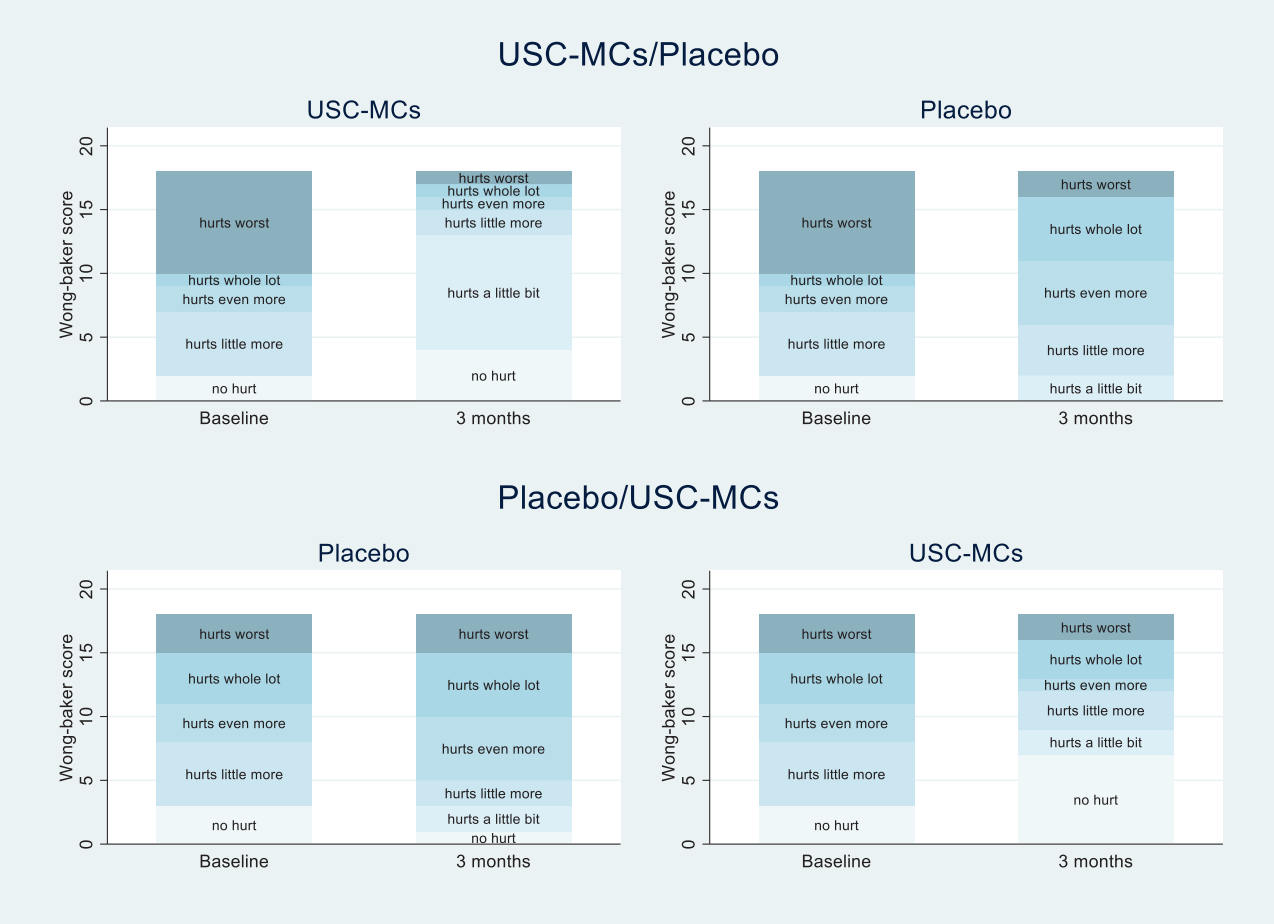

Figure 9. Wong Baker pain score at 3 months and period baseline, by sequence and period.

Table 12. Change in Wong Baker and VAS pain score, by sequence and period.

	Summary measures	Placebo/UC-MSCs (n=xx)	UC-MSCs/Placebo (n=xx)
Period 1	Mean (sd) Median (IQR) min, max	x.x (x.x) x.x (x.x, x.x) x, x	x.x (x.x) x.x (x.x, x.x) x, x
Period 2	Mean (sd) Median (IQR) min, max	x.x (x.x) x.x (x.x, x.x) x, x	x.x (x.x) x.x (x.x, x.x) x, x
Difference	Mean (sd) Median (IQR) min, max	x.x (x.x) x.x (x.x, x.x) x, x	x.x (x.x) x.x (x.x, x.x) x, x

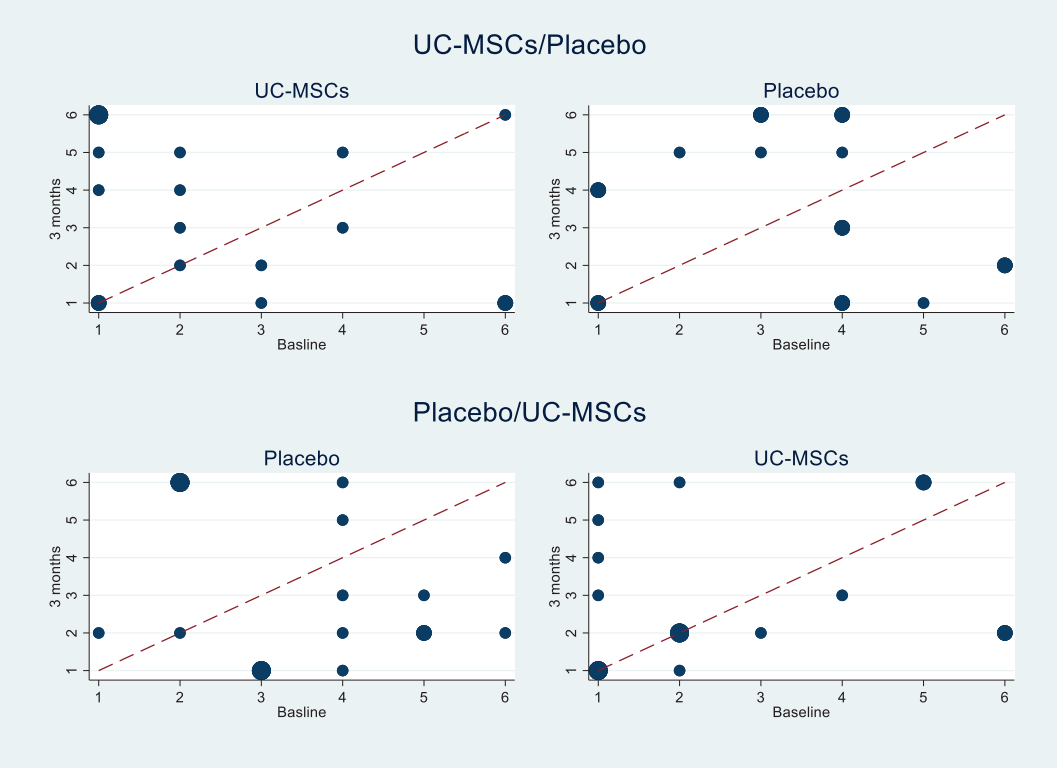

Figure 10. Scatter plot Wong Baker FACES scores at 3 months against period baseline by sequence and period.

15.13.2 Statistical analysis model

The effect of treatment on change in pain symptoms at 3 and 6 months for the Wong Baker FACES and VAS scales will be estimated based on mITT population using a mixed effects linear regression model as outlined for the EBDASI in Sections 15.10.3. If there are too few participants in the age range applicable to the Wong Baker scale the analysis will be descriptive only based on the discretion of the senior trial statistician during unblinded statistical analysis. Results will be presented as shown in Table 13.

Table 13. Effect of treatment on pain symptoms at 3 and 6 months post period baseline.

Secondary outcome	UC-MSCs/Placebo*		Placebo/UC-MSCs **		Paired adjusted mean difference in change (MSCs – Placebo) [95% CI] ǂ	Paired adjusted mean difference in change (MSCs – Placebo) [95% CI] †
3 months	N	LSM (95% CI)	N	LSM (95% CI)	No baseline adjustment	Baseline adjusted
Wong Baker FACES scores (⩾6 years)						
Change in average pain experienced in the last week	xx	xx(xx to xx)	xx	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
Change in worst pain experienced in the last week	xx	xx(xx to xx)	xx	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
VAS (All ages)	xx	xx(xx to xx)	xx	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
Change in child's average pain in the last week	xx	xx(xx to xx)	xx	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
Change in child's worst pain in the last week	xx	xx(xx to xx)	xx	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
6 months						
Wong Baker FACES scores (⩾6 years)	xx	xx(xx to xx)	xx	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
Change in average pain experienced in the last week	xx	xx(xx to xx)	xx	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
Change in worst pain experienced in the last week	xx	xx(xx to xx)	xx	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
VAS (All ages)	xx	xx(xx to xx)	xx	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
Change in child's average pain in the last week	xx	xx(xx to xx)	xx	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)
Change in child's worst pain in the last week	xx	xx(xx to xx)	xx	xx(xx to xx)	xx(xx to xx)	xx(xx to xx)

LSM = least squares mean

15.14 Analysis of general clinical appearance of wounds

This section covers the analysis of clinical assessment relating to the clinical assessment of changes in general wound appearance at 3 and 6 months compared to period baseline. We illustrate the analysis using changes in general wound appearance at 3 months post-infusion and the same approach when analysis general wound appearance at 6 months post-infusion. The three subdomains of the global score (arms legs and trunks) will not be analysed separately as the clinical research team are interested in the global score.

15.14.1 General considerations

A blinded clinician will assess wound photographs taken at months 3 and 6 post infusions and compare these to wound photographs taken at the start of each period (baseline). They will then make a clinical judgement on whether the observed change in wound appearance is classified as: a lot worse; a bit worse; much the same; a bit better or; much better.

A second clinician will independently provide a second opinion rating. Further details about data collection and derivation of the outcome, including the process if the 2 clinicians give different ratings, are provided in Section 18.8.

To aid interpretation of results, we will create and analyse an additional binary outcome relating to whether the observed changes in wound appearance showed any improvement.). For details of the analysis of these 5 levels (ordinal) and 2 levels (binary) outcomes see Sections 15.14.4 and 15.14.5, respectively.

15.14.2 Data visualisation and descriptive summaries

We will produce a stacked bar chart comparing the change in wound appearance by period and sequence as illustrated in Figure 11 and a scatter plot comparing the change in the treatment and placebo periods for each sequence as illustrated in Figure 12.

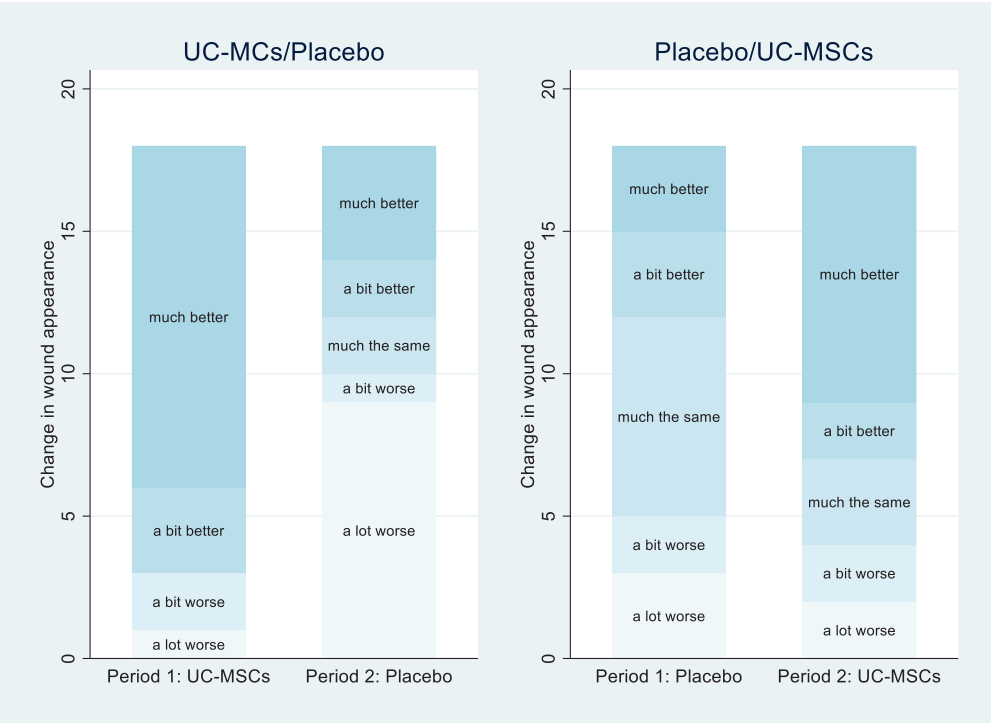

Figure 11. Change in clinical appearance of wound by sequence and period.

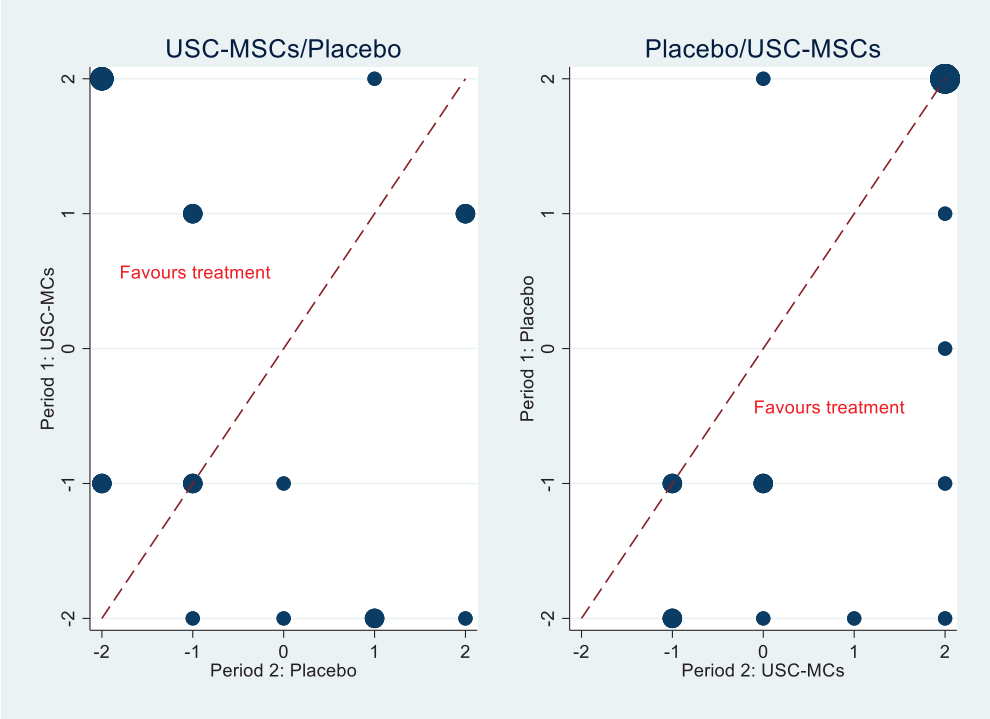

Figure 12. Scatter chart of change in clinical appearance of wound by sequence and period.

15.14.3 Discordance in treatment response between period within treatment sequence

For the binary outcome relating to whether there was an improvement in general wounds appearance of any amount, the within participant response to treatment will be summarised by treatment sequence and concordance or discordant pairs as shown in Table 14.

Table 14. Summary of any improvement in wound appearance.

Treatment sequence	Any improvement in wounds appearance in period (1, 2): (no, yes) = (0, 1)				Total
	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$	
Placebo/UC-MSCs	n_{11}	n_{12}	n_{13}	n_{14}	$n_{1\cdot}$
UC-MSCs/Placebo	n_{21}	n_{22}	n_{23}	n_{24}	$n_{2\cdot}$
Total	$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot 3}$	$n_{\cdot 4}$	$n_{\cdot\cdot}$

(a, b) indicates whether a participant experienced improvement in wounds appearance a in period 1 and b in period 2. n_{xy} $\{x, y = 1, 2\}$ represents the number of participants; for example, n_{11} is the number of participants in the placebo/UC-MSCs sequence who showed no improvement in both periods 1 and 2. As such, n_{12} and n_{21} (cells marked in green) indicate the number of participants who responded positively to UC-MSCs treatment in the placebo/UC-MSCs and UC-MSCs/placebo sequences, respectively. On the other hand, n_{13} and

n_{22} (cells marked in red) indicate the number of participants who responded better when a placebo was administered compared to UC-MSCs period in the placebo/UC-MSCs and UC-MSCs/placebo sequences, respectively. All participants in concordant cells marked in black (n_{11}, n_{14}, $n_{21}, and\ n_{24}$) do not convey any information about treatment effect as their responses were the same in the placebo and UC-MSCs periods in both sequences. Thus, the treatment effect is only informed by participants in discordant pairs in cells marked in green and red. For data visualisation, these summaries will also be displayed similar to Figure 12.

15.14.4 Analysis model for change in wounds appearance (ordinal outcome)

When necessary, the change in wound appearance (ordinal outcome) will be analysed using a mixed effects ordered/ordinal regression model with period and treatment as fixed factors and a random subject/participant effect term. This assumes a proportional log-odds meaning that the effect of treatment is identical at each cumulative category of the ordinal outcome. For an AB/BA crossover design, the mixed effects ordered logistic regression model can be mathematically characterised by Equation 3. This model is the same as a generalised mixed effects linear model of the ordinal family with a logit link function. The necessity of this model will be made at the discretion of the senior trial statistician depending on the observed distribution of responses across ordinal categories and the validity of the proportional log-odds assumption.

$$Y^*_{ijk} = b_0 + b_1 Period\ + \tau Treat\ + s_{ij} + \varepsilon_{ijk} \qquad \text{Equation 3}$$

Where:

- Y^*_{ijk} is a continuous latent variable such that y_{ijk}=-2 if $Y^*_{ijk} < k_1$; y_{ijk}=-1 if $k_1 \leq Y^*_{ijk} \leq k_2$, … and y_{ijk}=2 if $Y^*_{ijk} > k_4$;
- y_{ijk} is the change in Likert score between baseline and 3 months post baseline for subject i {i=1, … , n} of sequence j (j=1,2) in period k {k=1,2};
- b_0 is the intercept; $b_1$is the period effect; τ is the treatment effect averaged across the two treatment sequences;
- $s_{ij} \sim N(0,\ {\sigma_a}^2)$ is the subject/participant random effect which is assumed to be Normally distributed with mean 0 and variance ${\sigma_a}^2$ and;
- $\varepsilon_{ijk} \sim N(0, {\sigma_e}^2)$ is the residual error for the for subject/participant i of sequence j in period k and the residual errors are assumed to be Normally distributed with mean 0 and variance ${\sigma_e}^2$.

The maximum likelihood estimate (MLE) of model coefficients with SEs will be estimated using mean-variance adaptive Gauss-Hermite Quadrature integration. The treatment effect with 95%

CI will be reported in terms of the OR. To help readers interpret the results we will post estimate the marginal risks and summarise them with a stacked bar charts (Figure 13). This model can be fitted using the following STATA code.

```
// obtain treatment effect expressed as odds ratio
meologit outcome  i.trt  i.period ||id: , or
// summarise treatment effect using marginal predicted risks for each category
margins trt
// summarise treatment effect using marginal risk differences for each category
margins, dydx(trt)
```

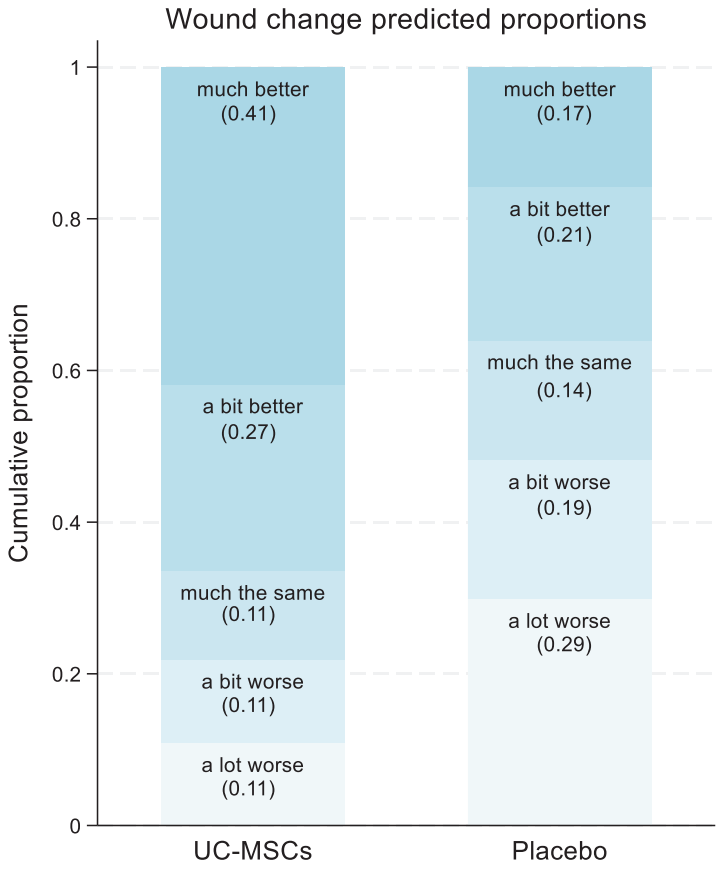

Figure 13. Predicted marginal risk of each wound change category by treatment.

15.14.5 Analysis of any observed improvement in wound appearance (binary outcome)

Irrespective of whether we model the outcome as an ordinal variable (Section 15.14.4), we will model the probability of seeing any improvement in wound appearance (a binary outcome following the definition based on the colour coded table in Section 15.14.3) using a mixed

effects logistic regression model. A binary outcome taking the value 1 if any improvement was observed (wound assessed as a little better or much better) or zero otherwise will be created. This "any improvement" outcome will be analysed using a mixed effects logistic regression model with treatment and period as fixed effects and a random subject/participant effect term. For an AB/BA crossover design, the mixed effects logistic regression model can be characterised by Equation 4.

$$(Logit\left(\pi_{ijk}\right) = b_0 + b_1 \text{Period} + \tau \text{Treat} + s_{ij} \qquad \text{Equation 4}$$

Where:

- π_{ijk} is the probability of observing an improvement for subject i {i=1, … , nj} of sequence j {j=1,2} in period k (k=1,2) at 3 months from period baseline;
- Logit $(\pi_{ijk}) = \log\left(\pi_{ijk}/\left(1-\pi_{ijk}\right)\right)$
- Y_{ijk} is an indicator variable of whether an improvement was observed or not for subject i {i=1, … , n} of sequence j {j=1,2} in period k (k=1,2) at 3 months from period baseline and Y ~B(1, π_{ijk});
- n is the number of people in each sequence; b_0 is the intercept; b_1 is the period effect; τ is the marginal treatment effect of interest averaged across the two treatment sequences and;
- $s_{ij} \sim N(0,\ {\sigma_a}^2)$ is the subject/participant random effect for participant i of sequence j, which is assumed to be Normally distributed with mean 0 and variance ${\sigma_a}^2$.

The estimated treatment effect expressed as an OR will be reported along with its associated 95% CI. RD and RR will also be presented to aid interpretation. The model can be fitted using the first line of the following Stata code when data are in long format and id is a participant indicator.

Marginal effect of treatment in terms of absolute RD will be post-estimated (e.g., using the Stata margins command). The estimated risk difference can be estimated using the second line of the Stata code in the box below. This code means the SEs of the estimated RDs are calculated using a truncated Taylor series (delta-method) whereas the SE of the OR is estimated using mean–variance adaptive Gauss–Hermite quadrature in the mixed effects logistic regression described above. To ensure consistency when reporting the 95% CI associated with the RD, we will use the z statistic for OR to calculate the 95% CI for the RD as RD ± 1.96/z × RD [19].

To estimate the marginal treatment effect and associated 95% CI in terms of approximate RR, a generalised mixed effects linear model of the Poisson family with a log link function and

robust SEs will be fitted with the same covariates included in Equation 4. This model will produce incidence RR that approximates RR. The model can be fitted using the final line of the Stata code in the box below.

```
// obtain treatment effect expressed as odds ratio
melogit outcome i.trt i.period ||id: , or
// obtain treatment effect expressed as risk difference
margins, dydx(trt)
// this is exactly as the above but using GLM framework
meglm outcome  i.trt i.period || id:, family(binomial) link(logit) or
// obtain treatment effect expressed as relative risk/risk ratio (incidence rate ratio)
meglm outcome i.trt i.period || id:, family(poisson) link(log) irr vce(robust)
```

15.15 Analysis of itch symptoms using the Itch man scale

Two self-reported questionnaire instruments are used for itch symptoms depending on the child's age. This section covers the Itch man scale for children aged 3 to 13. Section 15.16 covers the Leaven itch scale for children aged 14 and over. Self-reported itch symptoms are not collected for children under 3 but itch medication is recorded for all children (Section 15.17). The analysis is illustrated using the outcome at 3 months post infusion. The same approach will be used at 6 months post Infusion.

15.15.1 General Considerations

The Itch man scale is a child-reported ordinal outcome. Children are shown a series of five stick-men and asked which one best represents the extent of itching they are currently experiencing. Their choice of face is converted into a score between 0 (comfortable, no itch) to 4 (itches most terribly) with an increment of 1 between each. The clinical team has confirmed that there is an underlying (latent) continuum, and it is clinically meaningful to use the change in score associated with each category as a continuous outcome. The outcome (change in scores) will range from -4 to +4 with a negative score indicating an improvement in symptoms. Section 18.6 provides further details about data collection and derivation of the outcome. The same summary statistics as presented for the Wong-Baker FACES and VAS score will be presented for the Itch man (Section 15.13).

15.15.2 Analysis model

The analysis of the Itch man scale will follow the same approach as outlined for the Wong-Baker FACES and VAS score (Section 15.13.2). If there are too few participants in the age range applicable to the Itch Man scale the analysis will be descriptive only.

15.16 Analysis of itch symptoms using the Leuven itch scale

Two self-reported survey instruments are used for Itch symptoms. The choice of instrument depends on the child's age. This section covers the Leaven itch scale for children aged 14 and over. Section 15.15 covers the Itch man scale for children aged 3 to 13. Self-reported itch symptoms are not collected for children under 3 but itch medication data is collected for all children (Section 15.17). The analysis is illustrated using the outcome at 3 months post infusion. The same approach will be used at 6 months post infusion.

15.16.1 General Considerations

There are 12 domains in the Leaven itch. Six of these are of clinical interest. Itch severity, itch distress, and itch surface area are continuous outcomes with total scores between 0 and 100 and will be analysed using a similar approach to the EBDASI as outlined in Section 15.10.

Itch frequency, itch duration, and itch consequences are technically ordinal outcomes with a validated approach to convert them into a score between 0 and 100 provided by the instrument developers. The clinical team has confirmed it is clinically meaningful to use the change in score associated with each ordinal category as a continuous outcome. For consistency with the three continuous domains and results reported elsewhere we will use the converted scores between 0 and 100 as provided by instrument developers. All six outcomes (change in scores at 3 and 6 months from period baseline) will be in the range -100 to +100 with a negative score indicating an improvement in symptoms. Further details about data collection and derivation of the outcome are provided in Section 18.7.

15.16.2 Descriptive summary statistics and data visualisation

Summary statistics for the three continuous outcomes will follow the same approach as outlined for the EBDASI and iscorEB at 3 months post infusion as outlined in Section 15.10.2. Summary statistics for the three ordinal outcomes will follow the same approach as outlined for the Wong-Baker FACES and VAS scale at 3 months post infusion as outlined in Section 15.13.1.

15.16.3 Analysis of change in six Leuven itch domains

The analysis of change in scores in six itch domains of clinical interest at 3 and 6 months post period baseline (severity, distress, surface area, frequency, duration, and consequences) will follow the same approach outlined for EBDASI in Section 15.10.3. If there are too few participants in the age range applicable to the Leuven Itch scale the analysis will be descriptive only at the discretion of the senior trial statistician given the distribution of the data.

15.17 Analysis of changes in analgesia and itch medication

The section covers four outcomes at 3 months post infusion. These outcomes are not collected at 6 months post infusion.

1) change in regular pain medication,
2) change in as-required pain medication,
3) change in regular itch medication,
4) change in as-required itch medication.

We illustrate the approach using change in regular pain medication at 3 months after period baseline and the same approach will be used for the remaining three outcomes.

15.17.1 General considerations

Blinded clinician(s) will use the information on the case report form to compare medication taken by the participant in the last 48 hours at baseline and 3 months post infusion to makes a judgement about whether medication has:

1. reduced,
2. remained about the same or,
3. increased.

Clinical interest is focussed on a reduction in medication. As such, this will be analysed as a binary outcome taking the value 1 if clinicians judge that a reduction has occurred (1 above) and the value 0 if they judge that a reduction has not occurred (2 and 3 above). Sections 18.9 and 18.10 provide further details about data collection and derivation of the outcome.

15.17.2 Descriptive summary statistics and data visualisation

Although the analysis will focus on a binary outcome indicating whether a reduction has occurred, we will summarise all three outcomes, for example, as illustrated in the stacked bar chart in Figure 14. In addition, data will be presented similar to improvement in wounds appearance of any amount as detailed in Section 15.14.3.

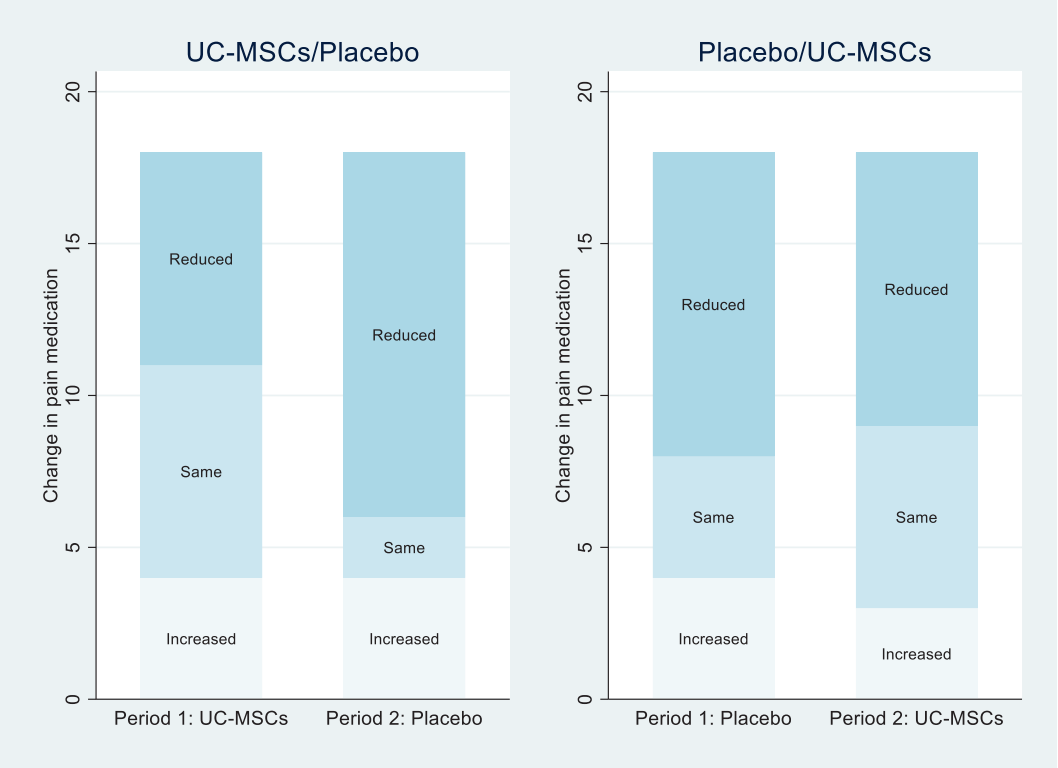

Figure 14. Change in regular pain medication at 3 months post-infusion.

15.17.3 Analysis of change in pain and itch medications (yes or no)

As described in Section 15.17.1, the four outcomes stated in Section 15.17 are binary variable taking the value 1 if medication has reduced and 0 otherwise. These will be analysed using the same approach outlined for any improvement in wounds appearance in Section 15.14.

15.18 Analysis of the CHU-9D Health utility

15.18.1 General Considerations

Child's health-related quality of life will be measured at baseline and 3 and 6 months post-baseline using the CHU-9D scale [13]. The CHU-9D is a sensitive and validated nine item assessment scale developed specifically with and for children and will be used in children aged 7 years and over. An age appropriate by proxy version will be completed by parents or caregivers for children aged 3 – 6. There is no measurement of health-related quality of life for children aged under 3. The questionnaire has 9 questions with 5 descriptive health-state response levels per question. A set of preference weights, giving utility values for each descriptive state are then applied [19]. Further details about data collection and applying the weights to produce utility values is derivation of the outcome are provided in Section 18.11.

15.18.2 Descriptive statistics and analysis

The statistical analysis will measure the change in health utility at months 3 and 6 months post baseline. Details of the derivation of the utility score are in Section 18.11. The same analysis methods and data presentation as outlined for the continuous outcomes in Sections 15.10.2 to 15.10.5 will be used and applied to the mITT population only (Section 12.1). The full health economic analysis using the CHU-9D is covered in a separate Health Economics Analysis Plan (HEAP).

15.19 Analysis of safety outcomes

15.19.1 General Considerations

The analysis of AEs and SAEs will be based on safety analysis population as defined in Section 12.5. This will focus on 4 categories:

- SUSARs within 48 hours of each infusion,
- SUSARs at any subsequent time after 48 hours of each infusion,
- SAEs (not exempted),
- AEs.

In addition, we will present subtypes within each category that will be agreed with the clinical team on a blinded basis during data freeze.

15.19.2 Descriptive statistics

For each of the 4 categories in Section 15.19.1, we will present descriptive statistics on the number and proportion of participants who reported at least one event of each category and at least one event of each subtype. We will also present descriptive statistics on the total number of events of each category and subtype and the total number of events of any category. In each case, numbers will be presented overall (i.e., across both sequences); by treatment arm (illustrated for SUSARs experienced within 48 hrs of any infusion in Table 15) and by period (treatment) within each sequence.

15.19.3 Data visualisation

Given the nature of the infusions, SAEs that occur closer to the infusion date are more likely indicative of potential harms than those that occur later in the follow-up period. As such, we will plot the frequency and type of SAE against time since first infusions by sequence and period (treatment) for all events that occur within 15 months of the first infusion. One option if to use a trellis plot (as illustrated in Figure 15) but the type of chart and the criterion for including an SAE will be determined with statistical discretion based on the pattern of the data we observe. For example, if an appreciable number of doses were missed, we might colour

code the SAEs by whether the dose was received. If there are clusters of SAEs of a particular type at a particular time, we might use size of the marker to represents the number of events.

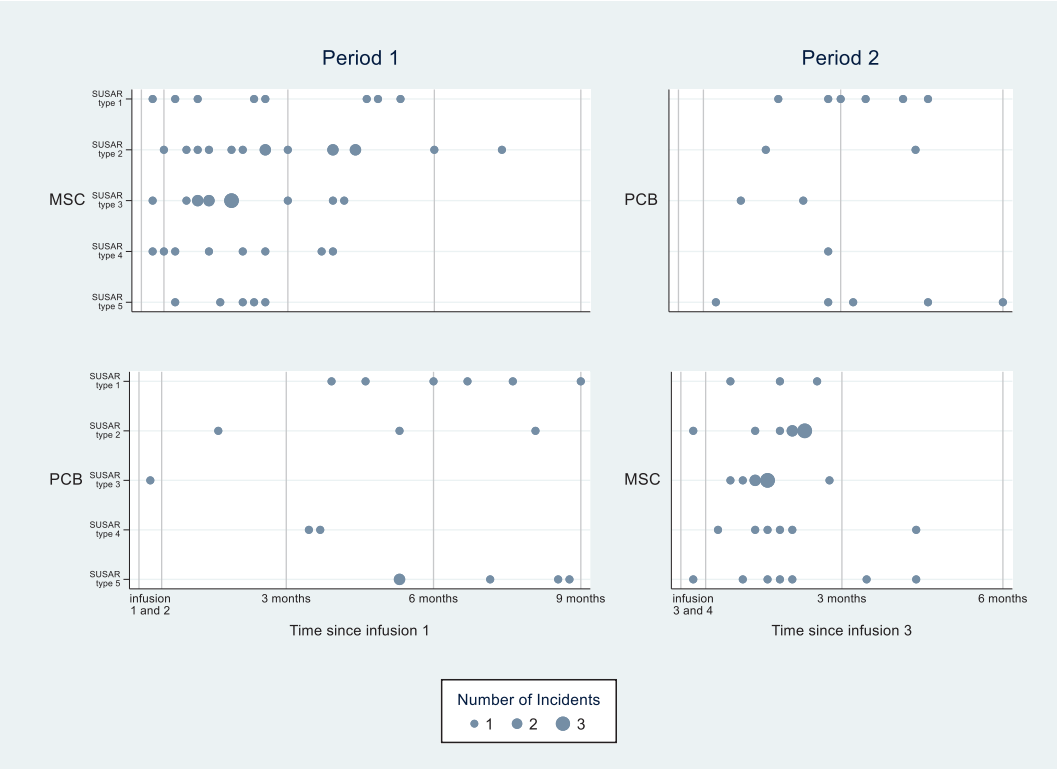

Figure 15. Trellis plot of SAEs.

15.19.4 Statistical analysis of adverse events

Additional formal statistical analysis will be performed in the event of marked repeated events and will account for differential follow-up of participants. This decision will be made at the discretion of the senior trial statistician depending on the observed distribution of the data. This will be done for the total number of events and number of repeated events of each type using a generalised linear mixed effects model of the Poisson family with a log link function with treatment and period as fixed effects and a random subject/participant effect term as well as follow-up time as an offset. Derivation of follow-up time is described in Section 18.14. Results will be presented as incident rates per arm and treatment effect between groups will be presented as incidence RR with a 95% CI. Section 17.3.5 describes model diagnostics and the alternative models we will consider if model assumptions are violated.

Table 15. SUSARs observed within 48 hrs of each infusion.

Classification	Details	UC-MSCs (N=x)						Placebo (N=x)					
		Total		Before 2nd infusion		After 2nd infusion		Total		Before 2nd infusion		After 2nd infusion	
		≥1 event (n)	All events (n)	≥1 event (n)	All events (n)	≥1 event (n)	All events (n)	≥1 event (n)	All events (n)	≥1 event (n)	All events (n)	≥1 event (n)	All events (n)
All SUSARs		Xx(%)	Xx	Xx(%)	xx	Xx(%)	xx	Xx(%)		Xx(%)		Xx(%)	xx
NCI CTCAE category	Blood and lymphatic system disorders		X		X		x						x
	...		...		...		...						...
	Vascular disorders		X		x		x						x
NCI CTCAE grade	Mild ...		X		X		X						X
	...		...		...		...						...
	Death related to AE		X		x		x						X
Relatedness to IMP	Reasonable possibility of being related		X		x		x						X
	...		...		...		...						...
	Not assessable		X		x		x						X
Relatedness to trial procedures	Reasonable possibility of being related		X		x		x						X
	...		...		...		...						...
	Not assessable		x		x		x						X
Related to another known cause	Yes		x		x		x						X
Seriousness	Death		x		x		x						X
	...		...		...		...						...
	Considered medically		x		x		x						x

	significant by the investigator												
Frequency	Isolated		x		x		x						x
	...		...		...		...						...
	Unknown		x		x		x						x
Intensity	Mild		x		x		x						x
	...		...		...		...						...
	Severe		x		x		x						x
Action taken	None		x		x		x						x
	...		...		...		...						...
	Other		x		x		x						x
Outcome	Recovered		x		x		x						x
	...		...		...		...						...
	Unknown		x		x		x						x

16 Summary of statistical analysis approach for each outcome

Table 16. Summary of outcomes and analysis approach.

Outcome (change in score)	Follow-up (months)	Analysis model(s)	Bayesian analysis	Analysis sets
EBDASI	3	Mixed effects linear regression	Yes	mITT, FA, CC, PP, MI
	6	Mixed effects linear regression	Yes	mITT, FA, CC, PP, MI
iscorEB	3	Mixed effects linear regression	Yes	mITT, FA, CC, PP, MI
	6	Mixed effects linear regression	Yes	mITT, FA, CC, PP, MI
Wounds appearance (ordinal)	3 and 6	Mixed effects ordered logistic regression; Mixed binary logistic (any improvement)	No	mITT
Any improvement in wounds appearance (binary)	3 and 6	Mixed effects logistic regression; Mixed effects Poisson regression	No	mITT
Wong Baker Faces (pain)	3 and 6	Mixed effects linear regression	No	mITT
Visual Analogue pain scale	3 and 6	Mixed effects linear regression	No	mITT
Itch man (ages 3-13)	3 and 6	Mixed effects linear regression	No	mITT
Leuven itch (ages 14+)	3 and 6	Mixed effects linear regression	No	mITT
Pain and itch medication (binary)	3 and 6	Mixed effects logistic regression; Mixed effects Poisson regression	No	mITT
CHU-9D	3 and 6	Mixed effects linear regression	No	mITT
AEs/SAEs	3 and 6	Descriptive; Mixed effects Poisson regression	No	mITT

17 Further details of statistical methods and calculations

17.1 Participant discontinuation

We anticipate minimal attrition. However, the analysis of participants who only complete one period needs careful consideration in a crossover trial. The inclusion or exclusion of participants who dropout following the first period can produce biased estimates in crossover trials depending on circumstances. The primary analysis is based on mITT that includes participants with outcome data in both periods as defined in Section 12.1. However, sensitivity

analysis will be performed using other analysis populations and analysis approaches thereby allowing us to explore the impact of discontinuations under different assumptions (Sections 12.112.2, 15.3 and 17.2.2).

17.2 Missing data and Imputation

17.2.1 Item non-response

For outcome measures with multiple questions in one or more domain we will impute up to 20% missing items in each domain deterministically using average response for the complete items in the domain.

17.2.2 Missing outcome data and Multiple Imputation (MI)

Multiple imputation will only be considered for EBDASI and iscorrEB at 3 months and 6 months post infusion using for the full analysis population, that is participants with at least one baseline period even if they are missing one or both period outcomes) (Section 12.2).

Under certain circumstances REML mixed models account well for missing data. As such, multiple imputation will only be undertaken at the discretion of the senior statistician based on investigation of descriptive summary statistics for missing data patterns and comparison of treatment effect estimates between analysis population. For each primary outcome we will summarise the frequency of missing data by reason (e.g., death, withdrawal, and lost to follow up) for each follow up visit overall and by period and sequence. We will also provide a summary of missing data patterns across visits (Table 17).

We will include all people with missing data in the MI including deaths but then combine the results in two ways, first by excluding deaths and second by including deaths.

Table 17. Missing data patterns for EBDASI and iscorrEB.

	Visit				Frequency		
	Infusion 1	Month 3	Infusion 3 (month 9)	Month 12	Overall (n)	UC-MSC/ Placebo (n)	Placebo / UC-MSC (n)
					n(%)	n(%)	n(%)
Pattern	1	1	1	1	x(x%)	x(x%)	x(x%)
	1	0	1	1			
	…	…	…	…			
	1	0	0	0			

1=data available, 0 =data missing

If considered appropriate we will multiply impute baseline and outcome scores using chained equations with age and RDEB type in the prediction model taking account of the longitudinal crossover nature of the design. Using data in long format (i.e., a separate row for each visit for each participant) we will impute any missing data for baseline, 3 months and 6 months post infusion using treatment, individual and RDEB type as discrete variables and age and timepoint as continuous variables. Only change between period baseline and 3 months post infusion will be used in the post-imputation estimation stage but incorporating the 6 months outcome in the MI takes better account of the available data. Treatment effect estimates using the primary analysis model (Section 15.10.3) will be pooled across 100 multiple imputations using Rubin's rules.

The procedure can be done using the below STATA code.

```
mi set mlong
mi register imputed iscoreb
mi register regular period treat individual_id time rdeb age
mi impute chained (regress) iscoreb = i.period time i.treat i.individual_id i.rdeb age, add(100)
burnin(1000)
mi passive : gen change = iscoreb - iscoreb[_n-1] if time==2 | time==5
mi estimate: mixed change i.period i.treat || individual_id: , reml cformat(%3.2f) pformat(%4.3f)
sformat(%5.3f)
```

* time takes the values 1 to 6 indicating: (1) baseline; (2) 3 months post infusion 1;(3) 6 months post infusion 1; (4) period 2 baseline/infusion 3; (5) 3 months post infusion 3 and (6) 6 months post infusion 3.

17.3 Checking statistical model assumptions

This section outlines our approach to model checking for each of the variable types discussed in Section 15.9. It should be noted that the sample size is relatively small as this is a rare condition, so the goal is not to achieve a perfect model fit.

17.3.1 Continuous outcomes

The analysis of continuous outcomes was covered in Section 15.10. Although the mixed effects linear regression model is relatively robust to non-normality assumption when analysing paired change scores, we will assess the normality assumption by exploring model residuals. The fitted values include the fixed linear predictor and a contribution from the person random

effect, with residuals calculated as usual. We will produce normal quantile plots and plots of residuals against fitted values. If there is evidence of extreme deviation from normality, we will use an appropriate transformation of the outcome as the primary analysis approach.

17.3.2 Ordinal outcomes treated as continuous

The mixed effects linear regression model is very robust to non-normality with paired change scores based on an underlying ordinal variable with a small number of categories. We will still investigate residuals using the approach outlined in Section 17.3.1 If assumptions are not met, appropriate transformation of the outcome data (e.g., log transformation) followed by back transformation of results to original scale of the outcome measure will be performed.

An issue more likely to cause problems with the model is observed outcome data which is not well spread over the full range of the scale. Taking the Wong-Baker score for example, if the change observed for each participant is either -2; 0 and 2, with no other outcomes observed, the proposed continuous model may not useful. In this eventuality, identifying an appropriate alternative will require statistician discretion. Options include collapsing the data into a smaller number of ordinal categories and using mixed effects ordered logistic regression (Section 15.14.4) or using mixed effects logistic regression with a binary variable indicating any improvement (Section 15.14.5).

17.3.3 Binary outcomes

17.3.3.1 Issues with zero cells

As there are only 4 possible outcomes for each sequence (Table 14) we may get zero participants in one or more cells. This 'zero cell' issue can lead to a statistical model not converging or unreliable estimates of treatment effect with extremely huge uncertainty. If we encounter missing cells and some of the coefficient estimates are considered unrealistically large, then at the discretion of the senior statistician we will report the outcome using descriptive statistics only.

17.3.3.2 Issues with model fit.

If the zero-cell issue does not occur, we will check the model in two ways.

We will check whether linear predictors are significant and squared linear predictor are non-significant predictors using Stata's '*linktest*' command. If this test is indicative of a specification error, identifying an appropriate alternative will require statistician discretion. Options include use of an alternative link functions and reporting the outcome using descriptive statistics only.

We will also calculate deviance and Pearson residuals and plot these against the linear predictors. If a data point for any participant is found to be highly influential, we will test the robustness of the estimated effect size by refitting the model with data for both periods from

that participant removed and report how this affects the treatment effect along-side details of the omitted participant (baseline characteristics, outcome status by period and SAEs).

17.3.4 Ordinal outcomes

For the ordinal model there is a greater risk of having no observations in one or more of the cells compared to the binary outcomes (Section 17.3.3.1) and in addition to this producing unrealistic model coefficients the estimation procedure may not converge. If the model does converge, the fit will be tested using Li & Sheppard [20] residuals which should be approximately uniformly distributed between plus and minus 1. Where possible, we will also assess the assumption of proportional odds directly by constructing a series of cumulative binary variables (level 1 v levels 2-5, levels 1-2 v levels 3-5 and so on) and estimating the OR for each. The 4 estimated ORs and their corresponding 95% CIs will be plotted (e.g., using a forest plot) to explore the validity of the assumption (i.e., homogeneity or consistency in the stratified ORs).

If any of these assumptions is violated, identifying an appropriate alternative will require statistician discretion. Options include collapsing the data into a smaller number of ordinal categories and using mixed effects ordered logistic regression (Section 15.14.4) or using mixed effects logistic regression with a binary variable indicating any improvement (Section 15.14.5). Any collapsing of the scale will be determined with clinical input blinded to the treatment allocation.

17.3.5 Adverse events (with time offset)

Although the generalised linear mixed effects model of the Poisson family with log link and robust SEs (Section 15.19.4) is generally robust to over-dispersion (i.e., the variance is appreciably larger than the mean) and variation in incidence rate we will investigate the extent of these deviations by plotting the distribution of the incidence rate (by period 1 and period 2). If the plot suggests an appreciable issue with over-dispersion, then at the discretion of the senior statistician we will consider using generalised negative binomial regression. We will also plot the number of events by time for each sequence (Section 15.19.3). If the between sequence difference in event rate is inconstant over time in either period (similar to non-proportional hazards) then also with the discretion of the senior statistician, we will consider weighting by time.

17.4 Bayesian model diagnostics and convergence

When using the MCMC sampling procedure, the reliability of Bayesian inference depends on whether the MCMC has converged. To check for MCMC convergence visually and formally, multiple chains will be used as described in Section 15.10.6. Visual inspection of graphical diagnostic plots will be conducted that include trace plots, the posterior distributions of parameters, and autocorrelation pots of all chains. For example, a clear separation of trace plots across chains or differences in posterior distributions between chains indicate MCMC non-convergence problem. Formally, based on these multiple chains, the maximum Gelman-Rubin (R_c) diagnostic criteria [18], [21] across all model parameters will be used to assess convergence. For example, a diagnostic R_c value above 1.2 is a non-convergence signal. In the case of non-convergence, alternative approaches will be used to improve convergence. For example, the burn-in period and the number of iterations can be increased, the MCMC sampling procedure changed, or the model specification amended.

18 Data manipulations and definitions

This section discussed the survey instruments and other data items collected in the trial; how they ae manipulated in preparation for analysis and how they are defined and interpreted. Section 17.1 outlines small differences in the data collected at each visit. The remainder of this section considers the individual survey instrument and data item.

18.1 General considerations

Once screening has taken place and consent has been taken there are 8 visits. Outcomes collected at each visit vary depending on the purpose of the visit as summarised in Table 18.

Table 18. Summary of purpose and timing of visits to hospital for treatment .

Period	Visit	Purpose (timing)	Variation in outcomes collected
1	1	Infusion 1 (day 0, period 1 baseline)	Excludes clinician judgement on changes in pain and itch medication.
	2	Infusion 2 (2 weeks after infusion 1)	Self-reported Pain and itch only (Wong baker, VAS, itch man and leaven itch)
	3	Follow up (3 months after infusion 1)	All outcomes collected
	4	Follow up (6 months after infusion 2)	All outcomes collected
2	5	Infusion 3 (9 months after infusion 1 and period 2 baseline)	Excludes clinician judgement on changes in pain and itch medication.

	6	Infusion 4 (2 weeks after infusion 3)	Self-reported Pain and itch only (Wong baker, VAS, itch man and leaven itch)
	7	Follow up (3 months after infusion 3)	All outcomes collected
	8	Follow up (6 months after infusion 3)	All outcomes collected

The pain and itch medications data collected in these booklets serve as the baseline against which clinicians assess changes in medication in visits 3, 4, 7 and 8.

18.2 The EBDASI

This outcome is completed by a clinician. The overall score [**range 0 to 506**] is the sum of the total activity score [**range 0 to 276**] and the total damage score [**range 0 to 230**]. Higher scores represent worse symptoms. The primary outcome is change since period baseline at 3 months post infusion. Lower change scores represent greater improvement. The scores are based on assessment of five anatomical locations as summarised in Table 19.

Table 19. Components of EBDASI score (0-506).

Anatomical Location	Activity	Damage
Skin [12 locations: ears, face, neck, chest, abdomen, back, buttocks, arms, hands, legs, feet, anogenital]	Each locations scored 0,1,2,3,5,7,8 or 10 for erosions, blisters, and crusting + the number of lesions (if <3) recorded but no impact on score. **[0-120]**	For each location, 7 types of damage scored (Erythema, Post-inflammatory hyperpigmentation or hypopigmentation, Poikiloderma Skin Atrophy, Hyperkeratosis, Scarring, Milia) as 0 (absence) or 1 (presence) **[0-84]**
Scalp	Scored 0, 1, 2, 3, 4, 8, 9 or 10 for severity of erosions and blisters + the number of lesions (if <3) recorded but no impact on score. **[0-10]**	Scored 0,1,2,3,4, 8, 9 or 10 for severity of 2 aspects: a) Post-inflammatory hyperpigmentation / Hypopigmentation OR erythema from resolving lesion OR hyperkeratosis and b) Scarring alopecia **[0-20]**
Mucous membranes	Scored 0,1,5 or 10 for severity of lesions at each of 12	For each of the 8 lesion types (ectropion, symblepharon, visible

	locations (Eyes, nose, buccal mucose, hard plate, soft palate, upper gingiva, lower gingiva, tongue, floor of mouth, labial mucosa, posterior pharynx, anogenital) + number of lesions at each location (if <3) recorded but no impact on score. **[0-120]**	corneal opacity, clinical microstomia, ankylogissia, intraoral scars, enamel hypoplasia, anal strictures) scored 0 or 2 for absence/presence **[0-16]**
Nails	Number with blistering, erosion, crusting and/or inflammation (0-5 for each of L/R and hand/foot) **[0-20]**	Number of nails with dystrophy/ calcification on hands/feet (0-20) + 2/3 × nails with Anonychia /related on hands / feet (0-40) **[0-60]**
Other epithelialized surfaces [5 locations: (larynx, oesophagus, genitourinary, hands and skin (cancer))	Only 3 used (larynx, oesophagus, genitourinary) and scored 0 (normal) or 2 (some activity) **[0-6]**	All 5 areas scored between 0 and 10 (number of levels used in each scale varies by location) **[0-50]**
Overall score	[**0-276**]	[**0-230**]

18.3 iscorEB

The iscorEB has a section completed by the clinician (Sections 18.3.1) and a section completed by the patient or parent/guardian (Section 18.3.2). The overall score [**range 0 to 258**] is the sum of the clinician score [**0-138**] and the patient score [**0-120**]. Higher scores represent worse symptoms. The outcome is change since period baseline at 3 months post infusion. Lower change scores represent greater improvement. The scores are based on assessment of five anatomical locations as summarised in Sections 18.3.1.

18.3.1 iscorEB questions completed by the clinician

The clinician section has five sub-scores as summarised in Table 20. Elements of iscorEB Clinician score [total range **0-138]:**

Table 20. Summary of iscorEB domains and scoring.

Score [Range]	**Description**

Skin involvement **[0-78]**	Four areas (head/neck, upper extremity, trunk and lower extremity) are given a score between 1 and 6 depending on extent of activity (from intact blisters to infections). The score for each area is then weighted, with different sets of weights for participants in different age bands (≤8 and >8). Finally, the weighted scores are multiplied by surface area affected (%). Score for the 4 areas are summed to give overall score.
Mucosal involvement **[0-15]**	Sum of separate score for each of 3 areas based on activity now or in the past 4 weeks. **Mouth score [0-2]** is score for erosions [0-1] plus score for opening distance between upper and lower incisors (compared to pre-specified percentiles) [0-2]. **Airway score [0-6]** is Stridor / hoarseness score (absence or number of days present) [0-3] plus EB related inhaled steroids use score (absence or number of days present) [0-3]. **Eye score [0-6]** is Eye redness / erosions score (absence or number of days present) [0-3] plus extent of Palpebral closure score [0-3]
Internal organ involvement **[0-12]**	Sum of separate score for 3 areas based on activity within the past 6 months. **GI / Nutrition [0-7]** is BMI score [0-4] (compared to pre-specified percentile ranges) plus extent of reliance on tube feeds [0-3]. **Urogenital score [0-3]** is based on renal disease [whether diagnosed and extent of treatment]. **Cardiac score [0-2]** relates to decreased cardiac function [diagnosis and presence of symptoms].
Laboratory abnormalities **[0-15]**	Sum of separate scores for 4 areas based on activity within the past 6 months. **Anaemia [0-4]** based on Hb level compared to pre-specified ranges. **Therapy for Anaemia [0-3]** (from none to transfusion). Albumin value **[0-3]** compared to pre-specified ranges. **Inflammation [0-5]** based on whether threshold values are exceeded for erythrocyte sedimentation rate (ESR); C-reactive protein (CRP) levels, platelet count (PLT) and serum Ferritin levels.
Complication / procedures **[0-18]**	Sum of separate scores for 4 areas based on activity within the past 6 months. **Squamous cell carcinoma [0-10].** [0=None; 1 new=1; SCC >2 new SCCs=2; nodal spread=5; metastatic disease=10]. **Osteopenia / osteoporosis [0-2].** [None=0; Normalized z-score

	≤2=1; non-traumatic fractures=2]. **Unscheduled hospital visits [0-3]** [None=0; EB-related emergency visit=1; EB-related emergency admission=2; EB-related ICU admission=3]. **Oesophageal dilation(s) [0-3]** based on frequency [None=0; 1-2=2; 3-4=3 ≥5=4].

18.3.2 iscorEB questions completed by the patient or parent/guardian

The patient or parent/guardian section has 15 questions grouped into seven categories as summarised in Table 21. Each question scored from 0 (none) to 8 (worst). The first part asks whether the form was completed by the participant or by the parent/guardian.

Table 21. Components of iscorEB Patient score (0-120).

Domain	**Questions**
Pain	5 items (overall, skin, mouth, eye, bone/joint) with faces and descriptions both shown, and each scored between 0 (none) to 8 (worst possible) [**0-40**]
Itching	Single item scored form 0 (non) to worst possible (8) [**0-8**]
Essential Functions	4 items (eating / drinking, bowel movements, urinating/voiding, sleep) scored 0 (no problems) to 8 (worst possible) [0-32]
Sleeping disturbances	Single item cored from 0 (none) to 8 (unable to sleep) [**0-8**]
Daily activity difficulty	2 items (moving around, using hands) scored from 0 (none) to 8 (unable to use / do anything) [**0-16**]
Mood	Single item scored from 0 (happy) to 8 (very unhappy) [0-8]
Impact	2 items (leisure, work/school/learning) scored from 0 (none) to 8 (unable to do anything) [**0-16**]

18.4 Wong Baker FACES score

Provides a single pain score for children over 6. The participant is shown 6 faces ranging from a happy face (no pain) on the left through to an unhappy face with tears (hurts worse) on the right with scores ranging from 0 to 10 in increments of 2. The participant is told they do not have to be crying to choose the 'worst pain' face.

18.5 Visual Analogue Scale for pain

The questionnaire shows a straight line marked '0 no pain' at left end and '10 worst pain' at the right end with no further numbers or markings. Parents/guardians place an X at a point on the line which indicates the perceived extent of the child's pain in the last week. The

omission of numbers or marking allows respondents to think more intuitively about their response rather than being distracted by choosing an actual number. Their response is converted into a score between 0 and 100 based on the distance of their mark from the left side of the line.

18.6 The itch man scale (ages 4 to 13)

This is a series of five similar pictures of a square containing a face with stick arms and legs. Looking from left to right the picture changes in three ways: the face becomes less happy, the position of the arms and legs become more erratic and there is an increasingly large number of clouds of dots posited on and around the image. There are scores and short descriptions below each picture ranging from 0 (comfortable, no itch) to 4 (itches most terribly; impossible to sit still; concentrate)

18.7 Leuven itch score (ages 14 and over)

This is a complex mixture of question types, and the manual does not provide a validated way to calculate an overall score. Given this complexity we will report each question separately. Each question is either a continuous outcome, an ordinal response converted into a continuous outcome, a nominal response without an associated continuous score or a free text field (see Table 22).

There are three domains based on an ordinal response:

- Itch frequency (0, 1, 2, 3, 4),
- Itch duration (0, 1, 2, 3),
- Itch consequence (11 questions, each scored 0, 1, 2, 3, 4).

There are three domains based on a continuous response (between 0 and 100):

- Itch severity (0 to 100),
- Itch distress (0 to 100),
- Itch consequence (11 questions, each scored 0, 1, 2, 3, 4).

The developer provides different methods for calculating a continuous outcome score.

- For itch frequency, the first step of the algorithm extends the range by taking the ordinal scores (0, 1, 2, 3) and transforming them to 0, 33 66 and 100 respectively. These scores are then averaged across groups of participants, e.g., the USC-MCs/Placebo group at 3 months post infusion.
- For itch duration, the first step is again to extend the ordinal range as above. A group score is then calculated by averaging the scores over only those participants with an itch frequency score greater than 0.

- For Itch consequence, the outcome is based on 11 questions about how frequently a participant's itching effects various aspects of their life, with responses between 0 (never) and 4 (always). Each of the 11 responses is converted from the values 0, 1, 2, 3 and 4 to 0, 25. 50, 75 and 100 respectively; they are summed and divided by 11 to give a score between 0 and 100. A group score is then calculated by averaging the overall scores from each participant with an itch frequency score greater than 0.
- For itch severity, the questionnaire shows a straight line marked 'no itch' at left end and 'worst possible itch' at the right end. Participants place an X at a point on the line which indicates the extent of their itching in the last month. Their response is converted into a score between 0 and 100 based on the distance of their mark from the left side of the line. A group score is then calculated using the subset of participant who scored greater than 0 on the itch frequency domain.
- For Itch distress, the questionnaire shows a straight line marked 'not distressing at all' at the left end and 'very distressing' at the right end. Participants place an X at a point on the line which indicates the extent to which their itching has been distressing in the past month. Their response is converted into a score between 0 and 100 based on the distance of their mark from the left side of the line. A group score is then calculated using the subset of participant who scored greater than 0 on the itch frequency domain.
- For Itch surface area, participants are shown outlines of a human body (front and back) and asked to shade the areas of the body where they have itched in the past month. For scoring purposes, the body is split into 31 areas (16 on the front and 15 on the back) with a percentage of surface area, or weight, associated with each, such that the total equals 100%. The participant's score is calculated by summing the weights associated with each of the body parts they have shaded. A group score is then calculated by averaging the overall scores from each participant with an itch frequency score greater than 0.

Details of the other six outcomes are provided in Table 22, which also provides a brief summary of all 12 outcomes.

Table 22. Summary of domains in the Leuven Itch Scale.

	Description	Response type	Outcome
1	The first question asks about the frequency of itching (never, rarely, sometimes often or	Ordinal	Itch frequency score

	always). If 'never' is ticked the patient is asked why itching did not occur / return and asked to stop.		Mean (0-100) and SD
2	how long itching lasted,	Ordinal	Itch duration score Mean (0-100), SD
3	when in the day itching occurred (multiple answers allowed)	Nominal	Nominal
4	Whether itching related to weather, activity and wound status (multiple answers allowed)	Nominal	Nominal
5	Patients asked to put a mark on a straight line drawn between 'no itch' and 'worst possible itch' to indicate severity.	Continuous (sliding scale)	Itch severity score Mean (0-100), SD
6	How itching was treated (no, ointment, medication, other) with free text fields (multiple answers allowed). There is no question 7.	Nominal	Nominal
8	Various consequences of itching (11 items covering physical, sleep, social) are rated on a 5 point scale (never, rarely, sometimes, often, always) with a free text box for other consequences as item 12.	Ordinal (multiple questions)	Itch consequences score. Mean (0-100), SD
9	Manner in which itching manifested itself from tickling through to burning with a free text box for other.	Ordinal	None specified in manual
10	Patients asked to put a mark on a straight line drawn between 'not distressing at all' and 'very distressing'.	Continuous (sliding scale)	Itch distress score. Mean (0-100), SD
11	Shows a front and back image of the body and patient asked to shade the parts of the body that itched.	Shading a picture	Itch surface area score. Mean (0-100), SD.
12	A free text box for other remarks	Free text	Descriptive

18.8 Clinical appearance

Photographs of wounds will be taken for the areas affected and assessed by a blinded panel with 2 or more independent experts.

Would appearance at 3 months post infusion will be compared with would appearance at period baseline and the panel will return one of 5 judgements for:

1 a lot worse,
2 a bit worse,
3 much the same,
4 a bit better,
5 much better.

Separate assessments will be made for

- Trunk (front and back),
- Arms and,
- Legs.

The images will then be subject to a 'global' assessment for overall improvement, using the same 5-point scale. Where any discrepancy exists between the assessments of the blinded independent EB experts, the scores will be reconciled as follows based on clinical advice:

- If the scores differ by 2 or more on the Likert scale the chief investigator will adjudicate,
- If the scores differ by just 1 on the Likert scale, we will take the higher score.

The identical process will be followed at 6 months post-infusion.

18.9 Dose and frequency of pain medication

At each infusion and assessment visit the names of all analgesia taken will be recorded on the analgesia log together with the dose (unit); frequency; method of administration; start date (or before trial) and end date. At all assessments except the period baselines (infusion 1 and infusion 3) the clinician will be asked to compare the current analgesia log with the analgesia log for the period baseline and record two clinical judgements.

- Has the patient taken their regular pain medication:
    1. More than usual,
    2. About the same as usual or,
    3. Less than usual?
- Has the patient taken their 'as required' (PRN) pain medication:
    1. More than usual,
    2. About the same as usual or,
    3. Less than usual.

The data for these outcomes comes from the medication log and two sources:

1. The name and dose of pain medication prescribed in the CRF,
2. Questions in the participant booklet.

A clinician uses the information on the CRF to make a judgement on whether the amount of medication taken has reduced between baseline and 3 months post infusion. A judgement using both sources is required in case either the regular or the 'as required' medication and dose has changed.

18.10 Dose and frequency of itch medication

At each infusion and assessment visit the names of itch medications taken will be recorded on the itch medication log together with the dose (unit); frequency; method of administration; start date (or before trial) and end date. At all assessments except the period baseline (infusion 1 and infusion 3) the clinician will be asked to compare the current itch medication log with the itch medication log for the period baseline and record two clinical judgements.

- Has the patient taken their regular itch medication:
    1. More than usual,
    2. About the same as usual or,
    3. Less than usual?
- Has the patient taken their 'as required' (PRN) itch medication:
    1. More than usual,
    2. About the same as usual or,
    3. Less than usual.

A clinician uses the information on the CRF to make a judgement on whether the amount of medication taken has reduced between baseline and 3 months post infusion. A judgement using both sources is required in case either the regular or the 'as required' medication and dose has changed.

18.11 CHU-9D

Likert responses (1-5) will be converted into utility scores using UK utility preference weights that have been provided by the developers. This conversion has been validated internally and implemented in the PROSPECT data management system. Scores are only calculated where all 9 questions have been answered (as recommended by the developers) [19].

18.12 Adverse events (AEs)

Monitoring of safety data by the DMEC including toxicities will continue throughout the trial. Reporting and recording will use the MedDRA dictionary using the standardised system organ classification (NCI CTCAE category) and any additional categorisation made by the clinical

investigators. Also reported is NCI CTCAE grade, relatedness to IMP, relatedness to trial procedures, seriousness, frequency, intensity, action taken and outcome.

18.13 Derivation of follow up date and per-protocol treatment windows.

Dates for follow up data collection and the start of period 2 are defined as 3, 6 and 9 calendar months since the first infusion on day 0 (irrespective of the months in between and the number of days in those months). Period 2 is analogous, with ideal planned follow at 3 and 6 calendar months after infusion 3. The per-protocol windows for each follow up point are provided in Section 12.4.

18.14 Calculating follow up period for safety analysis.

The statistical analysis comparing SAE frequencies between sequences will use the follow-up duration as an offset (Section 15.19.4). The length of follow-up will be derived for each period using the below algorithm to account for the washout period being in the first period only. The same algorithm will be applied when analysing all AEs. That is, if the final AE reported in the wash out period is an AE, the date of occurrence will still be used to determine the offset for the analysis of SAEs.

Period 1 participants:

- if a participant has no AEs or SAEs in the wash-out period, the follow up period will be the number of days between the date of the first infusion and the date of the 6 months visit physical exam [as recorded on the booklet completion form];
- if the participant has at least one AE or SAE in the wash-out period, the follow-up period will be the number of days between the date of the first Infusion and date of the last AE or SAE reported;
- if the participant discontinues, the follow-up period will be the number of days between the date of the first infusion and the date of discontinuation [as recorded on the participant completion and discontinuation form].

Period 2 participants:

- the follow up period will be the number of days between infusion 3 and the date of the last 6 months visit physical exam [as recorded on the participant completion and discontinuation form];
- if the participant discontinues, the follow-up period will be the number of days between the date of the third infusion and the date of discontinuation [as recorded on the participant completion and discontinuation form].

The follow up duration in each period will be calculated and normalised, in order to analyse incidence rate per person year of follow-up as follows:

$$follow\ up\ duration = \frac{((date\ of\ last\ follow\ up) - (date\ of\ day\ 0\ infusion))}{365.25}$$

AEs and SAEs collected after the participant has completed or discontinued [as recorded on the participant completion and discontinuation form] will be excluded from the analysis.

18.15 Summary of questionnaire instruments

Table 23. Summary of research instruments used in the study.

Name	Items	Range	Brief Description	Interpretation
Epidermolysis Bullosa Disease Activity and Scarring Index **(EBDASI)**	152	0-506	Number and severity of wounds and related activity, rated by clinician	Higher scores represent more severe symptoms
Instrument for scoring clinical outcomes of research for epidermolysis bullosa **(iscorEB)**	41	0-258	Similar to EBDASI, separate ratings by clinician (26) and patient or parent/caregiver (15)	Higher scores represent more severe symptoms
General clinical appearance	1	n/a	Blind independent assessment of change using photographs	1,2 = worse, 3 = no change, 4,5 = better
Wong Baker FACES pain score	1	0-10	Pictograms for Children over 6	Higher scores represent more pain
Visual analogue pain scale	2	0-10	Worst and average pain in last week	Higher scores represent more pain

Leuven Itch Score (full)	12	n/a	No overall score	higher scores generally mean more severe symptoms
Itch Man score	1	0-4	Pictograms for Children aged 4 to 13	higher scores generally mean more severe symptoms
Change in pain medication	n/a	0-1	Clinical assessment based on medicine, and dose from meds log and reported changes in frequency for regular and 'as needed' (PRN) meds	1 = decrease (better), 0= no decrease
Change in itch medication	n/a	0-1	As pain meds	As pain meds

19 Implementation of the SAP

This SAP will be used as a work description for the statistician involved in the trial. All analyses should ideally be performed by the trial statistician under the supervision of senior trial statistician. As part of quality control in line with the related Sheffield CTRU SOP, another CTRU trial statistician will perform independent programming and analysis of specific outcomes.

The DMEC will have access to unblinded data at their request during the trial; this data will be prepared by the data management team in the CTRU, aided by another CTRU statistician not involved in the trial when required.

Prior to database freeze the trial statisticians will be blind to outcome data at patient, treatment group and aggregate level. A CTRU statistician not otherwise involved in the trial will produce DMEC safety data reports during the internal dose de-escalation study.

The trial statisticians will be unblinded to produce the interim report (Section 13.2.4).

Following database freeze, the data manager will provide blinded patient level data for preliminary checks by the statistician. At an appropriate point between freeze and lock, un-blinded patient level data will be delivered to the statistician to define analysis sets and test statistical programs. Any queries will be communicated to the data manager prior to database

lock, and any changes to the database during this time will be documented. The database will be locked after agreement between the statistician, data manager and study manager. It is expected that no data amendments should be required following database lock. However, if an amendment is required, the process is documented in CTRU SOP DM012.

20 References

[1] S. V. Jain *et al.*, "The Epidermolysis Bullosa Disease Activity and Scarring Index (EBDASI): grading disease severity and assessing responsiveness to clinical change in epidermolysis bullosa," *J. Eur. Acad. Dermatology Venereol.*, vol. 31, no. 4, pp. 692–698, 2017, doi: 10.1111/jdv.13953.

[2] A. L. Bruckner *et al.*, "Reliability and validity of the instrument for scoring clinical outcomes of research for epidermolysis bullosa (iscorEB)," *Br. J. Dermatol.*, vol. 178, no. 5, pp. 1128–1134, 2018, doi: 10.1111/bjd.16350.

[3] G. Petrof *et al.*, "HHS Public Access," vol. 135, no. 9, pp. 2319–2321, 2015, doi: 10.1038/jid.2015.158.Potential.

[4] EU, "Appendix 4: Directive 2001/20/EC of the European Parliament and of the Council of 4 April 2001," *Textb. Pharm. Med.*, no. April 2001, pp. 827–843, 2007, doi: 10.1002/9780470987391.app4.

[5] ICH E9 Expert Working Group, "ICH Harmonised Tripartite Guideline. Statistical principles for clinical trials. International Conference on Harmonisation E9 Expert Working Group.," *Stat. Med.*, vol. 18, no. 15, pp. 1905–42, 1999, doi: 10.1002/(SICI)1097-0258(19990815)18:15<1903::AID-SIM188>3.0.CO;2-F.

[6] C. Gamble *et al.*, "Guidelines for the content of statistical analysis plans in clinical trials," *JAMA - J. Am. Med. Assoc.*, vol. 318, no. 23, pp. 2337–2343, 2017, doi: 10.1001/jama.2017.18556.

[7] K. Dwan, T. Li, D. G. Altman, and D. Elbourne, "CONSORT 2010 statement: Extension to randomised crossover trials," *BMJ*, vol. 366, 2019, doi: 10.1136/bmj.l4378.

[8] M. Dimairo *et al.*, "The Adaptive designs CONSORT Extension (ACE) statement: A checklist with explanation and elaboration guideline for reporting randomised trials that use an adaptive design," *BMJ*, vol. 369, pp. 1–34, 2020, doi: 10.1136/bmj.m115.

[9] V. Homer *et al.*, "Early phase clinical trials extension to guidelines for the content of statistical analysis plans," no. April 2018, pp. 1–10, 2022, doi: 10.1136/bmj-2021-068177.

[10] L. Zhao, S. Chen, P. Yang, and H. Cao, "The role of mesenchymal stem cells in

hematopoietic stem cell transplantation : prevention and treatment of graft-versus-host disease," vol. 9, pp. 1–13, 2019.

[11] C. Wong, D. and Baker, "Pain in children: comparison of assessment scales," *Nursing (Lond).*, vol. 14, no. 1, pp. 9–17, 1988.

[12] et al Haest C, Casaer MP, Daems A, "Measurement of itching: validation of the Leuven Itch Scale," *Burns*, vol. 37, no. 6, pp. 939–950, 2011.

[13] G. Furber and L. Segal, "The validity of the Child Health Utility instrument (CHU9D) as a routine outcome measure for use in child and adolescent mental health services," *Health Qual. Life Outcomes*, vol. 13, no. 1, pp. 1–14, 2015, doi: 10.1186/s12955-015-0218-4.

[14] C. Yap *et al.*, "Enhancing reporting quality and impact of early phase dose-finding clinical trials: CONSORT Dose-finding Extension (CONSORT-DEFINE) guidance," *Bmj*, 2023, doi: 10.1136/bmj-2023-076387.

[15] J. P. A. Ioannidis *et al.*, "Better reporting of harms in randomized trials: An extension of the CONSORT statement," *Ann. Intern. Med.*, vol. 141, no. 10, pp. 781–788, 2004, doi: 10.7326/0003-4819-141-10-200411160-00009.

[16] European Medicines Agency, "ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials - Step 2b," vol. 44, no. August, pp. 1–23, 2017.

[17] A. Gelman, D. B. Rubin, A. Gelman, and D. B. Rubin, "Inference from Iterative Simulation Using Multiple Sequences Linked references are available on JSTOR for this article : Inference from Iterative Simulation Using Multiple Sequences," *Stat. Sci.*, vol. 7, no. 4, pp. 457–472, 1992.

[18] S. P. Brooks and A. Gelman, "General methods for monitoring convergence of iterative simulations)?," *J. Comput. Graph. Stat.*, vol. 7, no. 4, pp. 434–455, 1998, doi: 10.1080/10618600.1998.10474787.

[19] K. Stevens, "Valuation of the Child Health Utility 9D Index," vol. 30, no. 8, pp. 729–747, 2012.

[20] C. Li and B. E. Shepherd, "A new residual for ordinal outcomes," no. March, pp. 473–480, 2012, doi: 10.1093/biomet/asr073.

[21] A. Gelman and D. B. Rubin, "Inference from Iterative Simulation Using Multiple Sequences," *Stat. Sci. 1992, Vol. 7, Pages 457-472*, vol. 7, no. 4, pp. 457–472, Nov. 1992, doi: 10.1214/SS/1177011136.

21 Appendix

Appendix 1 – Amendment History

SAP version	Date issued	Details of changes made
1	23/01/22	n/a
2	xx	• Clarification that participants not dosed in the DDI phase for logistic reasons will be excluded from the DDI analysis in section 11 • Membership of analysis populations in case of death, withdrawal, missed and late doses and timing of data collection added to section 12 • Presentation of non-compliance summary data added in section 15.6.2 • Descriptive tables of when infusions received and whether per-protocol added in section 15.8 • Complexities in the collection of iscorrEB data noted in sections 15.10.9 and 15.10.10 • Weakly informative priors for the Bayesian analysis added in section 15.10.6 • Procedure for dealing with discrepancies in wound appearance ratings is added in section 15.14.1 • Confidence interval calculation when estimating risk differences using the margins method added to Section 15.14.5. • Clarification that changes in pain and itch medication are not collected at 6 months post infusion in section 15.17. • Details of descriptive statistics and visualisations to be used for adverse events added in sections 15.19.2 and 15.19.3 • Multiple Imputation methods added to Section 17.2.2. • Model checking for discrete and adverse event outcomes added to Section 17.3 • Analysis of the Child Health Utility 9D (CHU-9D) added to sections 15.18 and 18.11. • Computation of follow up times for Per Protocol and Adverse event analysis added to sections 18.13 and 18.14. • The recommended dose of 2-3 million MSC cells per Kg added completion of the DDI phase (Section 12.4)
2-3		

Appendix 2

Table 24 Study procedures

Crossover Trial	Internal Phase 1, placebo-controlled randomised trial								
		Period 1[1]				Period 2[4]			
				9m washout before Period 2[2]					
Visit #	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Purpose	Screening	1st infusion Day 0 Period 1*	2nd infusion Day 0 Period 1 + 14d (+/- 5 days)	Month 3 (+/- 7 days)	Month 6 (+/- 14 days)	3rd infusion / Month 9 Day 0 Period 2[3] (+ 8 weeks)*[v]	4th infusion Day 0 Period 2 + 14d (+/- 5 days)	Month 12 (Period 2 Month 3) (+/- 7 days)	Month 15 (Period 2 Month 6) (+/- 14 days)
Patient information and informed consent	x								
Confirmation of consent	x	x	x	x	x	x	x	x	x
Inclusion/exclusion criteria review	x [†]								
Skin biopsy**	x								
Pregnancy test***	x	x	x	x	x	x	x	x	x
Physical examination	x	x	x	x	x	x	x	x	x
Vital signs	x	x	x			x	x		
Blood DNA analysis (if not done)	x								
IMP/Placebo infusion		x	x			x	x		
iscorEB		x		x	x	x		x	x
EBDASI	x	x		x	x	x		x	x
Pain assessment		x	x	x	x	x	x	x	x
Itch assessment		x	x	x	x	x	x	x	x
Photography of wounds		x	x	x	x	x	x	x	x
Changes to analgesic/itch medication				x				x	
Quality of life assessment		x		x	x	x		x	x
Routine bloods (safety bloods)****	x			x[+]		x		x[+]	
C7 serum antibodies (safety bloods)	x					x			
Serum cytokines (research bloods)		x	x			x	x		
Adverse event assessment		x	x	x	x	x	x	x	x
Concomittant medication assessment (including analgesia and itch)	x	x	x	x	x	x	x	x	x
Pain and itch medication taken 48 hours prior to study visit		x		x		x		x	

1 *Visit windows for Period 1 should be calculated from Period 1 Day 0;*

2 *The time from Day 0 period 1 (infusion 1) to Day 0 period 2 (infusion 3) should be at least 9 months*

3 *Visit windows for Period 2 should be calculated from Period 2 Day 0.*

ł Inclusion/Exclusion criteria must reviewed at screening and must be signed off by an appropriate member of the study team before a patient is randomised.
+ These routine bloods should only be taken if there are no clinical blood results available between the infusion date and follow-up visit. These routine bloods are not mandatory.
¥ Please note delays to the month 9 visit should be discussed and agreed with the central team. Ideally, this should be within 8 weeks of the due date.
** Note – outcome measures will be taken pre-dose*
*** For immunofluorescence. ONLY required if this information is not already in the patient notes. If needed, the skin biopsy will be performed according to local procedures.*
**** ONLY required in female participants who are menstruating and confirmed sexually active*
***** Blood tests: Full blood count, bone profile, liver function tests, renal profile, ferritin, CRP, ESR to be analysed in local NHS laboratories. The blood tests will be part of routine clinical care and no additional tests are required for the purpose of this study. The vital signs should be recorded every 15 minutes for two hours on the day of infusion. Refer to protocol section 6.2.1.*