TEEMS: Targeted treatment Early with Etanercept (biosimilar) plus Methotrexate or methotrexate with T2T care for DMARD-naïve early RA patients - Rationalising treatment based on naïve CD4+ T-cell Stratification.

STATISTICAL ANALYSIS PLAN

v3.0

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

Version date: 12/05/2025
Author(s): Shannon Halmkan, Rebecca Smith, Gill Parkinson, Kalpita Baird
Chief Investigator: Paul Emery
Trial Manager/Coordinator(s): Oksana Danyliuk

Contents

1. Document scope and relevant SOPs and guidance documents

This statistical analysis plan (SAP) was drafted prior to the completion of database lock and was prepared according to YTU SOPs and guidance documents.

2. Definition of terms/acronyms

Term	Definition
ACPA	Anti-citrullinated protein antibody
AE	Adverse event
CDAI	Clinical Disease Activity Index
CI	Confidence interval
CRF	Case Report Form
CRP	C-Reactive Protein
DMARD	Disease modifying anti-rheumatic drug
EMS	Early Morning Stiffness
ESR	Erythrocyte Sedimentation Rate
HAQ	Health Assessment Questionnaire
ICH	International Conference On Harmonisation
IMP	Investigational Medicinal Product
LFT	Liver function tests
LIRMM	Leeds Institute of Rheumatic and Musculoskeletal Medicine
MTX	Methotrexate
RA	Rheumatoid Arthritis
SAE	Serious Adverse Event
SD	Standard Deviation
SDAI	Simplified Disease Activity Index
SJC	Swollen Joint Count
T2T	Treat to target
TJC	Tender Joint Count
VAS	Visual Analogue Scale

3. Background and Rationale

The current optimal therapeutic approach in early Rheumatoid Arthritis (RA) is to start Methotrexate (MTX) to target inflammation and induce remission. Prediction of MTX therapy response remains a key clinical need to enable the identification of patients who would benefit from an alternative, more aggressive treatment strategy. Multiple predictors of remission with MTX have been reported over the years but none have entered routine clinical practice.

It has previously been reported by this study team that T-cell phenotyping at baseline could predict remission in disease modifying anti-rheumatic drug (DMARD)-naïve early RA treated with MTX. Reduced naïve CD4+ T cell frequency was the most predictive factor, using both a pilot and a replication cohort. The data confirmed the potential value of using naïve CD4+ T-cells as a biomarker of MTX induced remission in early RA. The clinical utility of measuring T-cell subsets is therefore strongly indicated by these data and suggests that measurement of T- cell subsets can be used to rationalise the use of MTX as first-line therapy.

Predicting response to MTX has important clinical value to identify patients who will do well on MTX, but furthermore for directing those who will have a sub-optimal response to MTX to receive alternative therapy without any harmful delay and in line with the treat to target (T2T) principle.

This study aims to validate the clinical value of T-cell subset quantification for the prediction of MTX response in early RA, by stratified interventions based on baseline naïve CD4+ T-cell status.

This study aims to assess the clinical utility of T-cell stratification for rationalising treatment with MTX, therefore enabling early targeted treatment for those with a poorer prognosis based on their immunological status.

4. Design

This study is a single centre, phase IV, open-label, prospective, three-armed longitudinal cohort study with an embedded pilot randomised controlled trial that aims to assess whether MTX can be rationalised as a first-line treatment for DMARD-naïve early RA patients, according to baseline naïve CD4+ T-cell stratification. The three treatment arms are given in Table 1.

Table 1: Trial arm descriptions

Arm	Description
Arm A	Methotrexate as per standard practice for patients with a normal T-Cell frequency
Arm B	Methotrexate as per standard practice for patients with abnormal T-cell frequencies
Arm C	Methotrexate in combination with subcutaneous Etanercept in patients with abnormal T-cell frequencies

The flow of participants through the study can be seen in Figure 1. Full details of the background and trial design are presented in the protocol (version 6.0).

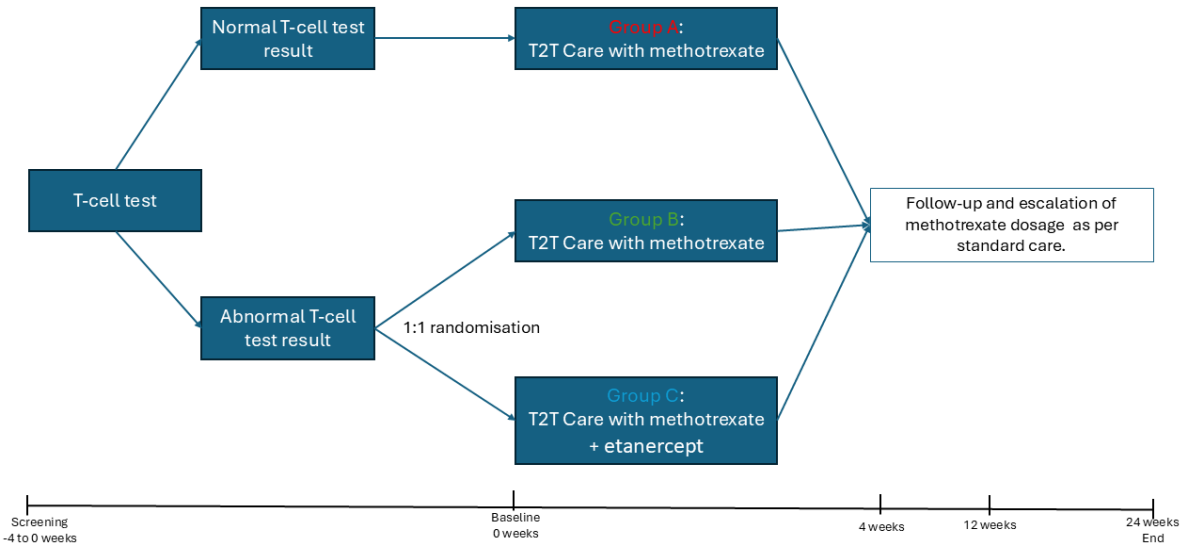


Figure 1: Schematic diagram showing study timeline and interventions

5. Trial Objectives

This study aims to assess the clinical utility of T-cell stratification for rationalising treatment with MTX, therefore enabling early targeted treatment for those with a poorer prognosis based on their immunological status. The study further aims to confirm and validate the clinical value of T-cell subset quantification for the prediction of MTX response in early RA, by stratified interventions based on baseline naïve CD4+ T- cell status.

The primary research hypothesis to be tested is whether a greater proportion of patients with normal naïve CD4+ T-cell frequencies for their age (Arm A) will achieve clinical remission after 24 weeks of first-line therapy with MTX and T2T care compared to those with abnormally low naïve CD4+ T-cell frequencies (Arm B).

5.1 Primary objective

To determine whether the proportion of patients achieving clinical remission after first-line treatment with MTX and T2T care differs according to naive CD4+ T-cell frequency (normal/abnormal) at baseline (Arm A vs. Arm B).

5.2 Secondary objectives

- To determine the proportion of patients that achieve clinical remission for patients with abnormal baseline naïve CD4+ T-cell frequencies receiving etanercept + MTX (Arm C) to obtain preliminary proof-of-concept of superior response compared to MTX with T2T care (Arm B).
- To determine whether the proportion of patients that achieve imaging remission after first-line treatment with MTX differs according to naïve CD4+ T-cell frequency (normal/abnormal) at baseline (Arm A vs. Arm B).
- To determine the values of patient-reported outcomes (EMS, VAS scales, HAQ-DI) for patients with abnormal baseline naïve CD4+ T-cell frequencies receiving etanercept + MTX (Arm C) to obtain preliminary proof-of-concept of superior response compared to MTX with T2T care (Arm B).
- To determine whether patient-reported outcomes (EMS, VAS scales, HAQ-DI) after first-line treatment with MTX differ according to naïve CD4+ T-cell frequency (normal/abnormal) at baseline (Arm A vs. Arm B).
- To determine the proportion of patients that achieve imaging remission for patients with abnormal baseline naïve CD4+ T-cell frequencies receiving etanercept + MTX (Arm C) to obtain preliminary proof-of-concept of superior response compared to MTX with T2T care (Arm B).
- To determine the proportion of patients who normalize their T-cell status for patients with abnormal baseline naïve CD4+ T-cells receiving MTX with T2T care or etanercept + MTX (Arm B & Arm C).
- To determine whether the proportion of patients achieving sustained clinical remission differs between study arms.
- To determine whether the cumulative amount of corticosteroid use differs between study arms.

6. Follow-up

An overview of data collection, and the corresponding time frames, is detailed below and summarised in Table 9.1 in the protocol. Data will be collected by clinical research networks at study recruitment sites and directly from participants. Laboratory samples will be tested by the NHS laboratory in the Leeds Teaching Hospital Trust and the results then entered onto the electronic database by a member of the research team. Electronic data records such as x-ray and ultrasound records will be saved and stored in an appropriately secure location as described in the Source Data Verification form.

The total duration of the study for patients completing all visits from screening to final assessment will be 28 weeks maximum (up to 4 weeks between screening and baseline visit).

6.1 Week -4 to 0: Screening/Consent Data

- Eligibility
- Consent
- Demographics
- Clinical observations
- Medical history
- Laboratory tests
- Joint counts

- CDAI
- SDAI
- DAS28
- HAQ-DI
- VAS
- X-rays

6.2 Week 0: Baseline/IMP Commencement

- Medication, allergy, co-morbidity check
- Joint counts
- CDAI
- SDAI
- DAS28
- HAQ-DI
- VAS
- Laboratory tests
- Ultrasound
- AE recording
- DMARD as T2T
- Etanercept for Arm C

6.3 Week 4: Safety Visit

- Compliance with treatment
- Joint counts
- CDAI
- SDAI
- DAS28
- HAQ-DI
- VAS
- Laboratory tests
- AE recording
- DMARD management

6.4 Week 12: Secondary Outcomes

- Compliance with treatment
- Joint counts
- CDAI
- SDAI
- DAS28
- HAQ-DI
- VAS
- Laboratory tests
- AE recording
- DMARD management

6.5 Week 24: Primary Outcome Timepoint

- Compliance with treatment
- Joint counts
- CDAI
- SDAI

- DAS28
- HAQ-DI
- VAS
- Laboratory tests
- AE recording
- DMARD management
- Ultrasound

6.6 Withdrawal Visit

- Change of status

6.7 Unscheduled Visit

- AE recording
- Changes to treatment to be documented
- Compliance with treatment
- Clinical observations
- Medication, allergy and co-morbidity check
- Clinical assessment data
- Joint counts
- DAS28
- CDAI
- SDAI
- Laboratory tests
- Patient reported outcomes
- VAS
- HAQ-DI
- Ultrasound
- DMARD Management

7. Outcomes

7.1 Primary outcome

The primary outcome is the difference in the proportions of patients in clinical remission at 24 weeks of first-line methotrexate with T2T care between those with normal (Arm A) vs. abnormal (Arm B) naïve CD4+ T-cell frequencies at baseline.

The Disease Activity Score (DAS28ESR) is a combined index that measures disease activity in patients with Rheumatoid Arthritis (RA). It has been extensively validated for its use in clinical trials. It incorporates the tender 28-joint count (TJC28), the swollen 28-joint count (SJC28), erythrocyte sedimentation rate (ESR; mm/hour) and patient global/general health assessment VAS (mm).

The TJC28 and SJC28 are calculated from the assessment of the same 28 joints where each joint is determined to be either tender or not tender and either swollen or not swollen. The assessment will be performed by an independent assessor to minimize bias. Where possible, the same assessor should preferably perform the TJC and SJC for a given patient particularly during the study to minimize inter-observer variation.

Missing joint assessments will typically result in a missing count, however there are several cases of this (see CTT20 ‘multiple IJA missed entry’**)** where they will be imputed as 0 due to confidence

expressed by the PI that the missing CRF fields should have indicated that the joint was not affected. In most of these cases only a few assessments were missing per participant.

The formula for DAS28(ESR) is:

$$DAS28(ESR) = 0.56\sqrt{TJC28} + 0.28\sqrt{SJC28} + 0.014(GH) + 0.70\ln(ESR)$$

The DAS28ESR will automatically be calculated within the MACRO system and computed at the analysis stage from the stored DAS28ESR components. Note that ESR values of 0 will be input as 0.001 to allow for them to be compatible with the DAS28 formula (i.e. ln(0) is undefined). The DAS28ESR ranges from 0-10 with higher scores indicating greater disease activity. If any of the components are missing, then DAS28ESR will be missing. Clinical remission (primary endpoint) will be deemed to have been achieved if DAS28ESR ≤2.6.

7.2 Secondary outcomes

7.2.1 Proportion of participants achieving clinical remission

The difference in the proportions of patients in clinical remission after 12 weeks of first-line methotrexate with T2T care will be calculated for comparisons between Arms A vs. B and Arms B vs. C. In addition, the difference in proportions of patients at 24 weeks will be calculated for Arm B vs Arm C only. Clinical remission will be deemed to have been achieved if DAS28ESR ≤2.6. Details of the DAS28ESR score are given in section 7.1.

7.2.2 Differences in the medians of PROMs at Weeks 12 and 24

The differences in the medians of patient-reported outcome measures (EMS, VAS scales, HAQ-DI) after 12 & 24 weeks will be calculated between Arms A vs. B and Arms B vs. C.

The EMS is the Assessment of early morning stiffness and is the duration of morning stiffness reported by subjects as the average daily length during the past week in minutes (from time of waking to time of maximal improvement, up to a maximum of 720 minutes).

The four VAS scales reported are:

- Physician Global Assessment of general health: a Visual Assessment Score from 0 – 100mm on the patient's current disease activity. The 0 mm corresponds to no disease activity, 100mm corresponds to extremely active.
- Patient Assessment of General State of Health: a VAS from 0 to 100mm, the patient rates their general state of health. 0mm corresponds to 'very well' and 100mm to 'very poor'.
- Patient Assessment of Disease Activity: a VAS from 0 to 100mm, the patient rates how their arthritis is, with 0mm corresponding to 'very well' and 100mm to 'very poor'.
- Patient Assessment of Pain: a VAS from 0-100mm, rating their current pain, with 0mm being no pain and 100mm corresponding to the worst possible pain.

The Health Assessment Questionnaire Disability Index (HAQ-DI) is formed of 20 questions in 8 categories of functioning that represent different activities - dressing, rising, eating, walking, hygiene, reach, grip and usual activities. For each item there is a 4-level difficulty scale scored from 0-3, representing no difficulty (0), some (1) or much (2) difficulty, and unable to do (3). The highest component score in each category determines the category score, unless the patient uses aids or devices for, or receives assistance with activities in that category, in which case the relevant category score is increased to 2 if the maximum score was previously <2. Where 'other aids' are

collected, a list of these will be produced and the category which they will contribute towards will be agreed and documented prior to database lock.

The score for the disability index is the mean of the eight category scores. If more than two of the categories, or 25%, are missing, the scale, cannot be scored. If fewer than 2 of the categories is missing, the sum of the categories will be divided by the number of answered categories. The higher score indicates greater disability.

7.2.3 Differences in the proportion of participants achieving imaging remission at Week 24

The difference in the proportions of patients in imaging remission (PD=0) at 24 weeks will be calculated between Arms A vs. B and Arms B vs. C.

Ultrasound assessment of power doppler synovitis (PD) will be performed at baseline and 24 weeks, as per standardised protocol on the standard set of 26 joints: Bilateral wrists, Metacarpal joints (MCP) 2 & 3, Proximal inter-phalangeal joints (2 & 3), Elbows, Knees, Ankles, Bilateral metatarsal joints (MTP) 1-5. Scans will be performed by the same dedicated ultra sonographer as per other clinical trials within the department at LIRMM.
PD will be scored between 0 and 3, with

- PD0 = no synovitis,
- PD1 = mild synovitis, ≤3 isolated single areas,
- PD2 = moderate synovitis, >3 isolated singles or confluent signal in less than half of the synovial area and
- PD3 = marked synovitis, signals in more than half of the synovial area.

Imaging remission will be determined if PD=0 (ie. No synovitis).

7.2.4 Difference in the proportion of participants achieving normalized T-cell status post-baseline

The difference in the proportion of patients with normal naïve CD4+ T-cells at 24 weeks for patients with abnormal baseline naïve CD4+ T-cells receiving MTX with T2T care (Arm B) vs. etanercept + MTX (Arm C) will be calculated.

Naïve CD4+ T-cells will be measured via cell surface antigen detection by 8-colour flow cytometry. The CD4+ T-cells will first be identified by staining for the CD and CD4 markers. Naïve cells will be identified based on the high expression of CD45RB and positivity for CD45RA and CD62L. Also included in this panel will be T-regulatory cells (Treg) and inflammation related cells (IRC's).

Patients' naïve T-cell frequency (%) will be compared to the level expected for their age according to the following formula derived from 106 healthy controls [1]: $(-0.54 * age) + 63.19$.

A patient whose naïve T-cell level is at or above this value will be considered to have a normal T-cell result; a patient with levels below the age-expected value will be considered to have an abnormal result.

7.2.5 Difference in the proportion of participants achieving sustained clinical remission between study arms

The difference in the proportions of patients in clinical remission (DAS28ESR ≤2.6) at both 12 and 24 weeks will be calculated for comparisons between Arms A vs. B and Arms B vs. C. The remission will be classified as sustained if the participant is in clinical remission at both time points. Details of DAS28ESR are given in section 7.1.

7.2.6 Difference in cumulative corticosteroid use up to and including Week 24 between study arms

The difference in the average (mean and median) cumulative amount of corticosteroid use at 24 weeks between study arms (Arm A vs. Arm B and Arm B vs. Arm C) will be calculated. This includes any administered at the 24 week visit also.

Patients in either study arm may receive corticosteroid treatment (IM/IA or oral) at any of the follow-up study visits or at unscheduled visits if they have evidence of clinical flare i.e. do not meet target DAS28ESR<2.6. Administered doses will be in line with local standard clinical practice. For patients meeting the target DAS score but that have evidence of flare i.e. synovitis as per the assessing physician, corticosteroids may be offered. The name, dose and type of corticosteroid will be logged in the Concomitant Medications CRF when prescribed. Dose equivalent tables [2] for different types of steroids will be used to convert doses relative to 5mg of Methylprednisolone.

The cumulative dose of corticosteroid used by each participant will be calculated by summing over each type of steroid prescribed at each visit, including unscheduled visits, at which steroids were prescribed during follow-up.

7.3 Other collected variables

At baseline, other variables collected include:

- Demographics: Age(years), gender, ethnicity
- Concomitant medications/allergy history,
- Surgical/medical history,
- Family history,
- Smoking history,
- Anti-citrullinated protein antibody (ACPA) titre and status (positive/negative),
- Rheumatoid factor antibody (RF) titre and status (positive/negative),
- Anti-nuclear antibody (ANA) titre and status (positive/negative),
- Date of RA diagnosis,
- Vital signs

SDAI

Simple disease activity index (SDAI) for RA is composed of the numerical sum of five outcome parameters: TJC and SJC (28 joints), patient and physician's global assessment of disease activity (range 0-10) and level of CRP (mg/dL). Classic disease activity cut-offs define low disease activity as an SDAI ≤11 and remission as a SDAI ≤5. Scores range between 0-100 with 100 indicating greater disease activity. The equation is:

$$SDAI = TJC28 + SJC28 + PGA + PhGA + CRP$$

It will be calculated at baseline and weeks 4, 12 and 24.

CDAI

The Clinical Disease Activity Index (CDAI) is derived from SDAI but does not include measurement of acute phase response. The computation of the score is done through the following equation:

$$CDAI = TJC28 + SJC28 + PGA + PhGA.$$

It will be calculated at baseline and weeks 4, 12 and 24.

Safety

The number of patients experiencing SAEs and AEs, any clinically significant worsening of a pre-existing condition, an AE occurring from overdose (accidental or intentional) of an IMP, an AE

occurring from abuse of an IMP or an AE associated with the discontinuation of the use of an IMP will be recorded.

Biochemistry and haemotology parameters will be collected at all visits.

Imaging parameters

Additional imaging parameters are collected at the same time as the PD including grey scale changes, erosions and osteophytes and tenosynovitis which are not covered in this SAP.

8. Data

8.1 Case Report Forms

The following CRFs are used in this study:

- Concomitant Medications
- Adverse Events
- Inclusion Criteria
- Exclusion Criteria
- Inclusion/Exclusion Criteria Met
- Informed Consent
- Randomisation and Study Meds
- Randomisation and Study Meds (wk4)
- Randomisation and Study Meds (wk12)
- Randomisation and Study Meds (wk 24)
- Changes in Medications
- Vital signs
- Medical History
- Physical Examination
- Screening Investigations
- Investigations
- T-cells
- TB testing
- Urine
- X-ray
- Examination – Systems
- Joint assessments
- Physician VAS
- Patient HAQ
- Patient VAS
- Disease Activity Scores
- Ultrasound

8.2 External datasets

Each subject's screening will be recorded in the Subject Screening Log. The Investigator will keep a list containing all subjects enrolled into the study. This list remains with the Investigator and is used for unambiguous identification of each subject. The list contains the subject identification number, full name, date informed consent signed, date of screening, and the hospital number or National Health Security number, if applicable.

The subject's consent and enrolment in the study must be recorded in the subject's medical record. These data should identify the study and document the dates of the subject's participation.

The laboratory measurements to be collected in this study include haematology, clinical chemistry, pregnancy testing (for females of childbearing potential), urinalysis, CRP, ESR and immunological measurements (T-cell subsets). The blood will be drawn and sent to the laboratory in accordance with standard NHS practice. T-cell samples will be sent for flow-cytometry. Sample labels containing appropriate identification information will be provided. In addition to the above, a maximum of 40 mL of blood will be drawn and stored for exploratory soluble biomarker testing at baseline and week 24. Approximately an additional 25 mL will be collected at visits 2-5.

Data from screening logs and laboratory measurements etc. needed for analyses will be contained in the MACRO system.

Protocol deviations
A list of the protocol deviations will be provided as a .csv file to YTU at the end of the study to be used in the analysis. Any significant deviations from the protocol that may influence the viability of the data will be flagged in the provided file.

8.3 Management of database and data verification

The research team is responsible for prompt reporting of accurate, complete, and legible data in the eCRF and in all required reports. Any change or correction to the electronic CRF will be time-stamped and documented within the data storage system. The Investigator will maintain a list of personnel authorized to enter data into the CRF.

Source documents are original records in which raw data are first recorded. These may include, but are not limited to, hospital/clinic/general practitioner records, charts, diaries, x-rays, laboratory results, printouts, pharmacy records, care records, or other printouts, completed scales, or Quality of Life Questionnaires. Source documents will be kept in a secure, limited access area.
Some data will be recorded directly in the CRF and will not appear in a physical source document (as defined in the Source Data Location Sheet document).
Source documents that are computer-generated and stored electronically will wherever possible be printed and filed in the patient medical notes for review by the monitor/inspector.

Source data will be transcribed into the eCRF, which also incorporates a validated electronic database, using a research data management system (MACRO). eCRF completion guideline documents are in place to highlight any data entry requirements and to minimise the need for queries. Computerized data cleaning checks will be used in addition to manual review to check for discrepancies and to ensure consistency of the data. Data will be entered into the research database using a rolling query-resolution system designed to identify data entry errors and protocol deviations in a timely fashion to allow accurate reconciliation.
An electronic audit trail system will be maintained within the MACRO system to track all data changes in the database once the data has been saved initially into the system or electronically loaded. Regular backups of the electronic data will be performed. A data export will be transferred to statisticians based in York trials unit via a University of Leeds OneDrive link.

9. Sample Size

We will allocate the patients to arm A if the patient's naive CD4+ T-cell frequency is normal for their age and gender. In our previous pilot study 42% of comparable RA patients had abnormal frequency values, of whom 25% responded to MTX at 6 months compared to 68% of those with normal values.

Those with abnormal values will be (randomly permuted block) randomised 1:1 to arm B or C, therefore this would result in an approximate 2.75:1 ratio of patients in arms A and B. For the primary outcome of clinical remission at 24 weeks, at the 5% level of significance, to achieve 80% power to detect a difference between the target proportions of patients achieving clinical remission in arms A (68%) and B (25%) [3], a total of 53 patients would be required, assuming binary logistic regression with a variance inflation factor of 1.05 estimated from pilot data. We will increase the sample size to n=60 to account for 10% drop-out. The total number needed to be recruited to the study would therefore be 76 (44 in arm A, 16 in arm B and 16 in arm C). To have a more robust estimate of the proportion of clinical remission in arm B and C, recruitment will continue until at least 34 and no more than 60 patients with abnormal T-cell values have been randomised 1:1 to arm B or C, and there are at least 46 patients in arm A. Thus, the total sample size for the study is anticipated to be between 76 (approximately A=44, B=16, c=16) and 106 patients (approximately A=46, B=30, C=30). In the case where recruitment within Arm A is disproportionately recruiting more than Arm B or C, recruitment will stop within Arm A once minimum recruitment number has been reached.

10. Randomisation and blinding

The pharmacy at Chapel Allerton Hospital will prepare a randomisation schedule. A central randomisation method using randomly permuted block sizes will be used. Patients will be randomised if they are found to have abnormal T-cells. The Trials team at Leeds General Infirmary will hold the master randomisation list and, as patients enter the study, they will be assigned the next available Patient ID Number and assigned to receive treatment with either MTX + Etanercept or MTX with T2T care. Randomisation codes will remain blinded to clinicians recruiting to the study as these will be administered centrally.

11.Final analysis

11.1 Analysis software

Analyses will be conducted at the end of the trial using Stata 18 *(StataCorp, 4905 Lakeway Drive, College Station, Texas 77845 USA*), or later (to be confirmed in final report) or R (version to be confirmed in final report).

11.2 Analysis principles and populations

For all analyses, unless otherwise stated, the analysis set will include all randomised participants. Analysis sets are given in Table 2.

Table 2: Analysis data sets

Analysis set	Description
Full analysis set (FAS)	The FAS will include all patients randomised and not withdrawn from the study prior to or during their baseline visit. Following the intent-to-treat principle, patients will be analysed according to the treatment assigned irrespective of treatment received.
Per protocol set (PPS)	The PPS will include patients in the FAS who have no significant deviations from the protocol that may influence the viability of the data.

Safety set	The Safety set will include all patients who took at least 1 dose of study treatment during the treatment period. Patients will be evaluated according to study treatment received.

The full analysis set will be used for all efficacy analyses unless specified otherwise and the safety set for safety analyses. Statistical tests for the primary and secondary outcomes comparing Arm A vs Arm B will be two-sided at the 5% significance level and 95% confidence intervals (CIs) shall be used. For the secondary outcomes comparing Arm B vs Arm C, no inferential tests will be carried out.

Due to the lower limit of reliable CRP detection being 5; values less than 5 will be recorded as 0 in the data. It will be assumed these values were uniformly distributed between 0 and 5 and hence will be replaced as 2.5 in the calculation of any scores (for example in calculating SDAI).

In general, summary statistics [n (number of available measurements), arithmetic mean, standard deviation, median, minimum, and maximum] for quantitative variables and absolute and relative frequency tables for qualitative data will be presented.

A table template to illustrate this with the primary outcome measure is given in appendix 1.

11.2.1 Hypothesis

The primary alternative hypothesis is that the difference (Δ) in the proportions of patients in clinical remission (DAS28 ≤2.6) after 6 months of first-line treatment with MTX monotherapy between those with normal (Arm A) vs. abnormal (Arm B) naive CD4+ T-cell frequencies will differ from 0.

$$H0: \Delta = 0$$
$$H1: \Delta \neq 0$$

The null hypothesis will be rejected if the p-value is less than 0.05.

11.2.2 Analysis visit windows

Visit windows will be used for all assessments unless specified otherwise. The actual day of an assessment will be mapped to the planned study visit using the following rules:

- If more than one assessment falls within a visit window, the closest non-missing assessment to the scheduled day will be used in the analysis.
- If more than one non-missing assessment actual dates are equidistant from the target day, the earlier assessment will be used in the analysis.

The windows for the analysis visits, split by the mid-point between the visits, are given in Table 3.

Table 3: Analysis visit windows

Timepoint	Study day	Analysis visit window in days
Week 0 (baseline)	1	All assessments prior to first dose of IMP
Week 4	29	Assessments post first dose of IMP to Day 57
Week 12	85	58 to 127
Week 24	169	≥ 128

The study day will be calculated as:

Study day = Actual visit or assessment date – date of initial IMP dose + 1

11.2.3 Missing data

Individuals who discontinue study treatment will continue to be followed up, where possible, according to the study visit schedule. Patients who discontinue study treatment early due to lack of efficacy will be considered treatment non-responders from that point onward (not achieving clinical

or imaging remission). In the main analyses, patients with missing data for reasons other than lack of efficacy will also be considered non-responders. For patient-reported outcomes, the analysis will only include patients with data available.

Missing data for the baseline covariates which will be used in analyses (DAS28ESR, age, gender, smoking status) will be imputed with the mean or the modal value (for categorical variables). If there is a large amount of missing data for a covariate then it will be considered whether to remove the variable from analysis models.

For the analyses of the T-cell status, missing age will be imputed with the mean for the calculation of the threshold for normal T-cell status. If the T-cell frequency is missing then the status will be set to abnormal if the missing data is due to lack of efficacy, otherwise the status will be set to missing.

In cases where there are a lot of missing data in secondary outcomes, best and worst-case imputation may be considered as ad-hoc analyses to assess the impact of exclusion.

11.3 Screening, eligibility, recruitment and follow-up data

The following information from the screening and eligibility processes will be reported in tabular format (by treatment arm and overall) and in the CONSORT flow diagram:

- The number of participants screened
- The number/proportion of eligible/ineligible participants
- Reasons for ineligibility
- The number of eligible participants approached for consent
- The number of participants that did/did not consent
- Reasons for non-consent

Frequency of and timing of withdrawals will be presented overall and by randomised group, with reasons where available. A template of the CONSORT diagram can be found in Appendix 2.

A recruitment graph will present the overall recruitment by month and this will also be available in tabular form. A graph of the actual vs target recruitment may also be produced where the target recruitment will be calculated based on the overall average recruitment rate per site per month.

The rate of missing data for each outcome at each follow up point will be presented by trial arm and overall, as stated in section 11.5.

11.4 Baseline data

Baseline data for all randomised participants (except any ineligible patients randomised in error) and for only those included in the primary analysis, will be summarised descriptively by trial arm and overall. Parameters summarised will include demographics, vital signs, smoking status, alcohol intake and medical history. Continuous baseline data will be summarised in terms of the non-missing sample size, arithmetic mean, standard deviation, median, interquartile range, minimum and maximum. Categorical baseline data will be summarised in terms of frequencies and proportions. No formal between group comparisons of baseline data will be undertaken. Tables illustrating this are given in Appendix 3.

Unless otherwise specified, the last valid measurement before study medication administration will be utilised as the baseline value.

11.5 Analysis of primary outcome

The primary outcome will be summarised descriptively by trial arm and overall at each time point, including the level of missing data.

11.5.1 Primary analysis

The estimand for the primary outcome is the difference in the proportion of patients in clinical remission (DAS28ESR ≤2.6) at 24 weeks of first-line methotrexate with T2T care between those with normal (Arm A) vs. abnormal (Arm B) naïve CD4+ T-cell frequencies, with patients who discontinue study treatment early due to lack of efficacy considered as treatment non-responders.
The primary endpoint will be analysed using binary logistic regression in the FAS. The model will adjust for DAS28ESR at baseline, age (at baseline), gender, smoking status (binary; current or previous vs never), and trial arm. Participants with missing data will be imputed as treatment non-responders.

The treatment effect will be presented in the form of both unadjusted and adjusted odds ratios, along with the corresponding two-sided 95% Wald CIs and p-values. Model coefficients for the covariates with 95% Wald CIs will additionally be presented to aid understanding of the fitted model. The model coefficients will be used to calculate the marginal proportions and difference in proportions. Two-sided bias-corrected 95% CIs will be calculated via non-parametric bootstrapping.

Model assumptions will be checked as follows: the multicollinearity between variables can be checked using the variance inflation factor (VIF), which measures the correlation and strength of correlation between the predictor variables in a regression model. Checks for influential observations and outliers will be conducted using Cook's Distance and if any are detected then sensitivity analyses will be completed to assess the impact on the model when these observations are removed. Graphs will be plotted of each explanatory variable against the logit of the response variable to assess if the relationship between them is linear. Transformations of the variables included in the model may be considered. If assumptions are not met, appropriate non-parametric tests will be considered.

If the model does not converge then smoking status will be removed from the model. If there are still issues with convergence, resulting in a sparse dataset and complete/ quasi-separation, then Firth logistic regression [4] may be used instead. This is similar to the logistic regression model but includes a penalisation term in the likelihood function.

Table templates for the primary outcome descriptive summaries and results are given in appendix 1.

The primary analysis will be checked by a second statistician on completion which will be documented using *F16: Primary Analysis Sign Off* Form.

11.5.1.1 Sensitivity analyses

Per protocol

The primary analysis will be repeated in the set of patients treated according to protocol (PPS).

Missing data

Details concerning the extent and reasons for missing primary endpoint data will be reported by trial arm. The primary analysis assumes that all patients with missing data are non-responders. This assumption will be tested by conducting sensitivity analyses using the FAS, including best-case single imputation for these patients (which corresponds to imputing them as responders), complete case analysis and, in addition, multiple imputation may be considered.

For participants with missing data due to lack of efficacy this will remain imputed as treatment non-responders.

Missing joint assessments

Several deviations have occurred indicating missing joint assessments and in the main model these were assumed as 0. Since these will impact directly on the primary outcome (see section 7.1), a sensitivity analysis will be conducted on the main primary model replacing these assessments as missing. This will result in missing DAS-28 scores and as per section 11.2.3, they will be imputed as treatment non-responders.

11.5.1.2 Supplementary analyses

Repeated measures logistic regression

To account for longitudinal data, a repeated measures logistic regression analysis maybe carried out if sample size and the number of treatment responders allows for convergence. The model in section 11.5.1 will be used with the addition of week and week-by-trial arm interaction as fixed effects and participant as a random effect. All available data will be used in the analysis with DAS28ESR ≤2.6 at a visit considered as response at that visit. Missing outcome data will be imputed as treatment non-responders. If there are convergence issues then using the QR decomposition of the variance-components matrix will be considered.

The conditional odds ratio (conditional on the covariates and assuming have the same random effect) for each visit along with 95% CIs will be reported. The marginal proportions and difference in marginal proportions will be calculated from the model coefficients. Two-sided bias-corrected 95% CIs will be calculated via non-parametric bootstrapping.

Change from baseline to Week 24 in DAS28ESR score

An additional analysis will be carried out using the DAS28ESR as a continuous measure instead of converting to a binary response, since information is lost when a variable is dichotomised.

The continuous DAS28ESR score will be summarised by visit, overall and by treatment arm.

The continuous DAS28ESR score will be analysed using a repeated measures mixed-effects linear regression model, with the change from baseline in DAS28ESR score at Weeks 4, 12 and 24 as the dependent variable, adjusting for baseline DAS28ESR, trial arm, age (at baseline), gender, smoking status (binary; current or previous vs never), week and trial arm-by-week interaction as fixed effects and including the participant as a random effect. An unstructured covariance matrix will be used to describe the correlations between DAS28ESR and visits for a participant. If this model fails to converge then other covariance matrix structures will be considered. The denominator degrees of freedom will be calculated using the Kenward-Roger method. All available data will be used in the analysis and no imputation will be made for missing data. If the model assumptions are not met, then transformations of variables or alternative regression models may be used.

The marginal mean estimate, mean difference between trial arms and associated Wald 95% CIs and p-values will be reported for each time point.

11.5.1.3 Subgroup analyses

There are no planned subgroup analyses.

11.6 Analysis of secondary outcomes

The secondary outcomes will each be summarised descriptively by trial arm and overall at each time point, including the level of missing data.

11.6.1 Proportion of participants achieving clinical remission at Week 12 between Arms A and B

This outcome will be analysed and reported as outlined for the primary outcome in Section 11.5.1, with the outcome switched to the proportion of participants achieving clinical remission (DAS28ESR≤ 2.6) at Week 12 in Arms A and B as opposed to Week 24.

11.6.2 Proportion of participants achieving clinical remission at Weeks 12 and 24 between Arms B and C

The difference in proportion achieving clinical remission at Weeks 12 and 24 in Arm B vs Arm C will be analysed as described in section 11.5.1. The proportions, unadjusted and adjusted odds ratios with a range of CIs (75%, 85%, 90% and 95%) will be presented.

11.6.3 Differences in the medians of PROMs at Weeks 12 and 24 between Arms A and B

Patient-reported outcomes (EMS, each VAS, HAQ-DI) at Weeks 12 and 24 will be summarised descriptively between arms A and B, means and standard deviations or medians and interquartile range (if skewed) will be presented.

11.6.4 Differences in the medians of PROMs at Weeks 12 and 24 between Arms B and C

As described in section 11.6.3 descriptive statistics will also be provided for Arms B vs C.

11.6.5 Differences in the proportion of participants achieving imaging remission at Week 24 between Arms A and B

The proportions of patients in imaging remission (PD=0) at Week 24 will be summarised descriptively by arm and overall for Arms A and B.

11.6.6 Difference in the proportion of participants achieving imaging remission at Week 24 between Arms B and C

As described in section 11.6.5, descriptive summaries will also be provided for Arm C.

11.6.7 Difference in the proportion of participants achieving normalized T-cell status post-baseline

The difference in the proportion of patients in Arm B vs. Arm C will be summarised descriptively by arm and overall.

11.6.8 Difference in the proportion of participants achieving sustained clinical remission between study arms

This endpoint will be assessed for Arms A vs B and again for Arms B vs C.

The model detailed in the primary analysis (section 11.5.1) will be utilised where the response variable is replaced with a binary indicator of sustained clinical remission. The proportions achieving sustained clinical remission and unadjusted and adjusted odds ratios with corresponding 2-sided 95% CIs will be reported for Arms A vs B and Arms B vs C. In addition, the p-value will be reported for Arms A vs B and additional CIs (75%, 85% and 90%) will be reported for Arms B vs C.

11.6.9 Difference in cumulative corticosteroid use up to and including Week 24 between study arms

Descriptive summaries of the corticosteroid dose at each visit by arm will be provided. The difference in the unadjusted average (mean and median) cumulative amount of corticosteroid use up to and including at the 24 weeks visit between study arms (Arm A vs. Arm B and Arm B vs. Arm C) will also be included.

11.7 Safety analyses

11.7.1 Adverse events

Adverse events for this trial will be summarised descriptively using the safety set. The number of serious and non-serious adverse events over the 24-week follow-up period, as well as the number and percentage of participants these events occurred in, and the exposure by 100 patient-years will be reported for each trial group.

The number of serious adverse events in each trial group will be broken down by reason for classification as serious adverse event and relationship of event to treatment and outcome. The number of non-serious adverse events in each trial group will also be broken down by relationship of non-serious adverse event to treatment and outcome.

11.7.2 Laboratory measurements

Laboratory parameters from haemotology (full blood count) and biochemistry tests (eGFR, LFT, urea and electrolytes) will be summarised descriptively by visit for each trial group and overall.

12 Exploratory analyses

12.1 Optimal naïve CD4 T-cell frequency cut-off

Previously published research suggested that the optimal naive CD4 T-cell frequency cut-off for prediction of response to MTX monotherapy was equivalent to the expected value for the patient's age. We will verify this by pooling the patients in arms A and B and using non-parametric ROC curve analysis, taking clinical remission at 6 months as the reference variable and age-corrected baseline naive CD4± T-cell frequency as the classification variable. Area under the ROC curve and the 95% confidence interval will be presented. The point at which the Youden Index (J = sensitivity + specificity – 1) is maximised will be chosen as the optimal cut-off. A plot of the baseline CD4 T-cell against the continuous DAS28ESR index at 6 months will be provided. We may additionally investigate whether there is evidence that baseline DAS28ESR affects the accuracy of classification by including it as a covariate in a secondary ROC analysis.

12.2 Other efficacy outcomes

Descriptive summaries of the other efficacy outcomes, (SDAI and CDAI), will be presented by follow-up visit and trial arm.

12.3 Protocol deviations

The number and percent of protocol deviations will be reported by arm. Protocol deviations which may influence the viability of the data (and result in the patient being excluded from the PPS) will be listed.

12.4 Change from baseline in DAS28ESR score

The change from baseline in DAS28ESR score will be considered in arms A vs B at 12 weeks and in arms B vs C at 12 and 24 weeks. The model specified in the latter section of 11.5.1.2 will be repeated but considering the relevant combinations of timepoint and trial arm comparisons.

13 SAP amendment log

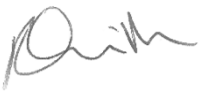
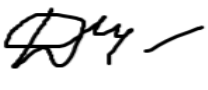
Amendment/addition to SAP and reason for change	New version number, name and date
SAP completed and signed-off	*V1.0, S.H. R.S. , 30/01/2025*
Section 7.1 and 11.5.1.1- Addition that missing joint assessments listed in CTT20 'multiple IJA missed entry' will be imputed as 0 in the main analysis models due to the confidence of the PI that these are most likely 0 and the large number of participants this impacts. Section 13.1: Changes to protocol analyses added	*V2.0, S.H. R.S. , 12/05/2025*
Section 11.2 Changes to definition of FAS added Section 13.1: Deviation from protocol definition of FAS added	*V3.0, S.H. R.S. , 21/05/2025*

13.1 Changes to protocol planned analysis

The following changes were made to the protocol planned analysis due to the statistician at Leeds leaving and the analysis being taken over by the YTU:

- Protocol section 12.5.2: Secondary analyses of imaging and PROMs outcomes removed, only descriptive statistics will be produced
- Protocol section 12.5.3 and 12.7: Sensitivity analyses updated to only be carried out for the primary outcome
- Protocol section 12.5.3: Supplementary analyses added to include a repeated measures logistic regression model and considering the change in continuous DAS28 score from baseline via a repeated measures linear regression model.
- Protocol section 12.4 Full analysis set to include all randomised patients who were not withdrawn from the study prior to or during their baseline visit rather than just randomised.

14 Signatures of approval

Name	**Trial Role**	**Signature**	**Date**
Shannon Halmkan	Statistician		21/05/2025
Rebecca Smith	Statistician		21/05/2025
Oksana Danyliuk	Trial Coordinator		21/05/2025

Paul Emery	Chief Investigator		21/05/2025

15 References

[1] Hunt L, H.E., Nam J, et al., T cell subsets: an immunological biomarker to predict progression to clinical arthritis in ACPA-positive individuals. Ann Rheum Dis, 2015.

[2] NICE. (2024). *Equivalent anti-inflammatory Doses of Oral Corticosteroids*. [online] Available at: https://cks.nice.org.uk/topics/corticosteroids-oral/background-information/equivalent-anti-inflammatory-doses/ [Accessed 22 Oct. 2024].

[3] Smolen, J.S., et al., A simplified disease activity index for rheumatoid arthritis for use in clinical practice. Rheumatology (Oxford), 2003. 42(2): p. 244-57.

[4] DAVID FIRTH, Bias reduction of maximum likelihood estimates, *Biometrika*, Volume 80, Issue 1, March 1993, Pages 27–38

[5] Koenker, R., and V. d'Orey (1987): “Computing Regression Quantiles”, Applied Statistics, 36, 383-393

16 Appendices

Appendix 1: Primary outcome table shells

Descriptive statistics of DAS28(ESR) and clinical remission by visit

	Group A (N=XX)				Group B (N=XX)				Group C (N=XX)			
	Baseline	**4 weeks**	**12 weeks**	**24 weeks**	**Baseline**	**4 weeks**	**12 weeks**	**24 weeks**	**Baseline**	**4 weeks**	**12 weeks**	**24 weeks**
Disease Activity Score-28 (DAS-28 (ESR))												
n												
Mean (SD)												
Median (IQR)												
Min-Max												
Missing												
Clinical Remission												
Clinical remission achieved (DAS28 ≤2.6) N (%)												
Clinical remission not achieved (DAS28 ≤2.6) N(%)												
Missing												
Sustained clinical remission achieved N(%)												

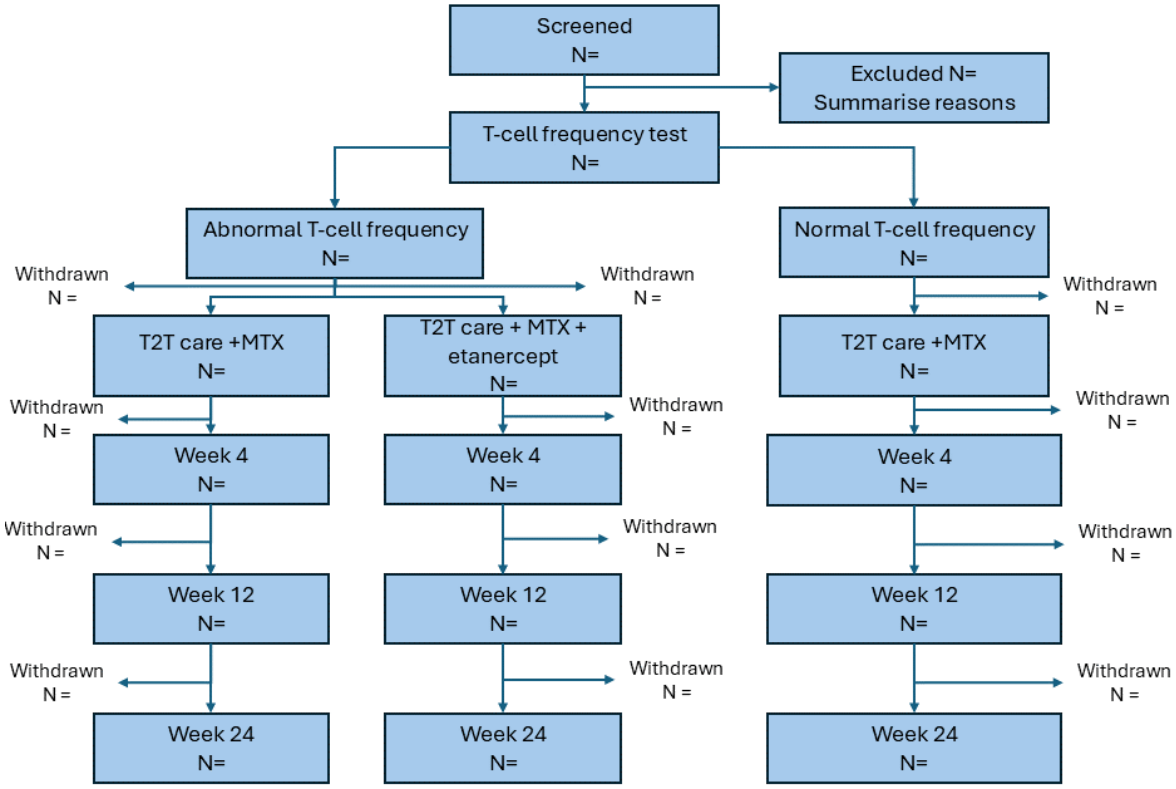
Similar descriptive statistics tables will be produced for PROMs, imaging remission, T-cell status, SDAI and CDAI and laboratory measurements.

Difference in proportions achieving clinical remission by visit, Full analysis set

	n	Proportion	Adjusted difference in proportions (95% CI)	Unadjusted odds ratio (95% CI)	Adjusted odds ratio (95% CI)	p-value
Week 4						
Arm A						
Arm B						
Week 12						
Arm A						
Arm B						
Week 24						
Arm A						
Arm B						

Trial Arms are defined as: Group A -T2T care with methotrexate (normal T-cell frequency), Group B-T2T care with methotrexate (abnormal T-cell frequency
Comparisons are relative to Arm B
Patients who discontinue study treatment early are considered as non-responders

Appendix 1: Trial CONSORT diagram



Appendix 2: Baseline Table Shells

Note: the following will be summarised for all participants as well as those only included in the primary analysis.

Baseline Demographics and characteristics table Shell

Baseline Demographics and characteristics				
	[1]Group A	[1]Group B	[1]Group C	Total
Age				
N				
Mean (SD)				
Median (IQR)				
Min-Max				
Missing				
Gender (N [%)				
Female				
Male				
Missing				
Ethnicity (N [%				
White				
Asian or Asian British				
Black, Black British, Caribbean, or African				
Mixed or multiple ethnic groups				

Prefer not to say				
Other[2]				
Missing				
Family History of Auto Immune Disease (N [%])				
Yes				
No				
Unknown				
Missing				
Smoking Status (N [%])				
Current				
Previous				
Never				
Missing				
Alcohol intake (units/week)				
N				
Mean (SD)				
Median (IQR)				
Min-Max				
Missing				

[1] Trial Arms are defined as: Group A -T2T care with methotrexate (normal T-cell frequency), Group B-T2T care with methotrexate (abnormal T-cell frequency), Group C- T2T care with methotrexate + etanercept (abnormal T-cell frequency)

Vital Signs/observations at Baseline

Vital Signs/observations at Baseline				
	[1]Group A	**[1]Group B**	**[1]Group C**	**Total**
Pulse (/min)				
N				
Mean (SD)				
Median (IQR)				
Min-Max				
Missing				
Systolic Blood Pressure (mmHg)				
N				
Mean (SD)				
Median (IQR)				
Min-Max				
Missing				
Diastolic Pressure (mmHg)				
N				
Mean (SD)				
Median (IQR)				
Min-Max				
Missing				
Temperature (°C)				
N				
Mean (SD)				
Median (IQR)				
Min-Max				

Missing					
Weight (kg)					
N					
Mean (SD)					
Median (IQR)					
Min-Max					
Height (m)					
N					
Mean (SD)					
Median (IQR)					
Min-Max					
Missing					
BMI (kg/m^2)					
N					
Mean (SD)					
Median (IQR)					
Min-Max					
Missing					
BMI Severity (N [%])					
Underweight (<18.5)					
Healthy Weight (18.5-24.9)					
Overweight (25-29.9)					
Obese (30-39.9)					
Severely Obese (≥40)					
Missing					
Duration of Electrical Muscle Simulation (mins)					
N					
Mean (SD)					
Median (IQR)					
Min-Max					
Missing					
*Other Clinically Significant diseases (N [%])					
Hypertension					
Hypercholesterolaemia					
Ischaemic heart disease					
Cerebrovascular disease					
Peripheral vascular disease					
Peptic ulcer disease					
Renal disease					
Chronic liver disease					
Epilepsy					
Demyelination					
Asthma					
Emphysema/Chronic bronchitis					
Diabetes**					
Depression					
Thyroid dysfunction					
Cancer***					

***Note categories are not mutually exclusive**

**** Of which X% are tablet controlled(x1% x2% x3% in arms A,B & C respectively), Y% are diet controlled (y1% y2% y3% in arms A, B & C respectively) and Z% are insulin dependent (z1% z2% z3% in arms A,B & C respectively).**

*****Cancer sites will be summarised here.**