Dose Reduction and Discontinuation of Prednisolone Using Structured Taper in Patients with Polymyalgia Rheumatica: Statistical Analysis Plan for a Randomised, Open-Label, Parallel-Group, Multi-Centre Trial

Trial Registration

ClinicalTrials.gov registration ID: NCT05636501

SAP version: 1.0 Date: 2025-06-02

Protocol version: This document has been written based on information contained in the study protocol version 1.2, dated: 08.10.2022

Roles and Responsibilities

SAP authors: Christoffer Våben, Robin Christensen, Tobias Haugegaard and Kresten Keller

Principal Investigator: Christoffer Våben, MD

Senior Biostatistician Responsible: Robin Christensen, BSc, MSc, PhD

Junior Biostatistician: Tobias Haugegaard, BSc, MSc

Sponsor: Kresten Keller, MD, PhD

Signatures

Name	Role	Title and Affiliation	Signature	Date
Christoffer Våben	Principal Investigator	PhD Fellow Department of Rheumatology, Aarhus University Hospital, Aarhus, Denmark		20/06 2025
Robin Christensen	Senior Biostatistician Responsible	Professor of Biostatistics and Clinical Epidemiology Section for Biostatistics and Evidence-Based Research, the Parker Institute, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark & Department of Clinical Research, University of Southern Denmark, Odense University Hospital, Denmark		2025, June 06
Tobias Haugegaard	Junior Biostatistician	Research Fellow Section for Biostatistics and Evidence-Based Research, the Parker Institute, Bispebjerg and Frederiksberg Hospital, Copenhagen, Denmark		2025, June 06
Kresten Keller	Sponsor	Associate Professor Department of Rheumatology, Aarhus University Hospital, Aarhus Denmark		20 June 2025

1. Introduction

1.1 Background

Polymyalgia rheumatica (PMR) is the most common inflammatory rheumatic disease among the elderly population (1). It is characterised by proximal symmetrical joint pain, morning stiffness, fatigue, and raised inflammatory markers. Prednisolone is the anchor drug in the treatment of PMR, but should be tapered and discontinued within the first 1-2 years of treatment to mitigate adverse effects such as infections, osteoporosis, and diabetes (2). However, prednisolone tapering can be challenging, and some patients need prolonged treatment (3).

Currently, PMR is diagnosed and treated in both primary and secondary care, depending on health care availability as well as local guidelines. The International European League Against Rheumatism (EULAR)/American College of Rheumatology (ACR) 2015 management guidelines also state that PMR can be managed in both primary and secondary care, but no studies have compared the two treatment approaches (2).

A key goal in PMR management is to taper prednisolone as fast as possible to reduce adverse effects, but at the same time keep the disease in remission. A retrospective study showed that when prednisolone tapering was carried out by nurses at the department of rheumatology following a structured tapering scheme and fixed guidelines for when patients should be assessed, patients reached a significantly lower daily dose of prednisolone one year after diagnosis and significantly more patients were in prednisolone free remission after 2 years, compared with a historical cohort without a specified structured taper regimen (4). However, the results need to be confirmed in a prospective setting.

The aim of this trial was to compare a nurse-led structured prednisolone taper in a hospital setting with usual care prednisolone tapering in general practice in newly diagnosed PMR patients.

1.2 Objectives

Primary Objective:

- To compare the effect of a structured taper strategy, relative to usual care, on the proportion of patients in prednisolone-free remission after 52 weeks, in patients with PMR.

Key Secondary Objectives:

- To compare the effect of a structured taper strategy, relative to usual care, on changes in the patient-reported prednisolone dose from baseline to week 52, in patients with PMR.
- To compare the effect of a structured taper strategy, relative to usual care, on number of relapses from baseline to week 52, in patients with PMR.
- To compare the effect of a structured taper strategy, relative to usual care, on the proportion of patients diagnosed with giant cell arteritis (GCA) during 52 weeks, in patients with PMR.
- To compare the effect of a structured taper strategy, relative to usual care, on changes in the patient-reported global VAS from baseline to week 52, in patients with PMR.

Secondary Objectives:

- To compare the effect of a structured taper strategy, relative to usual care, on changes in PMR-Activity Score (PMR-AS) from baseline to week 52, in patients with PMR.
- To compare the effect of a structured taper strategy, relative to usual care, on the proportion of patients with undiagnosed GCA verified by ultrasound at the 52-week visit, in patients with PMR.
- To compare the effect of a structured strategy, relative to usual care, on changes in SF-36 MCS from baseline to week 52, in patients with PMR.
- To compare the effect of a structured taper strategy, relative to usual care, on changes in SF-36 PCS from baseline to week 52, in patients with PMR
- To compare the effect of a structured taper strategy, relative to usual care, on changes in HAQ-DI from baseline to week 52, in patients with PMR.

- To compare the effect of a structured taper strategy, relative to usual care, on changes in PMR VAS from baseline to week 52, in patients with PMR.
- To compare the effect of a structured taper strategy, relative to usual care, on changes in fatigue VAS from baseline to week 52, in patients with PMR.
- To compare the effect of a structured taper strategy, relative to usual care, on changes in stiffness VAS from baseline to week 52, in patients with PMR.
- To compare the effect of a structured taper strategy, relative to usual care, on changes in morning stiffness duration (min) from baseline to week 52, in patients with PMR.

1.3 Hypothesis

The main hypothesis is that a structured tapering strategy is more effective at bringing patients into prednisolone-free remission after one year. In addition, it is expected that the patients still receiving prednisolone treatment will have reached a lower daily dosage after one year. Consequently, prednisolone-related side effects will be less frequent in the structured taper group. It is not expected that a difference will be seen in patient-reported outcome measures between the two groups, nor is it expected that a difference will be seen in the PMR-AS. It is expected that more patients will be diagnosed with GCA during one year in the structured tapering group compared with usual care tapering.

2. Study Methods

2.1 Trial Design

This study was designed as a multi-centre, randomised, parallel-group, open-label trial. Newly diagnosed, treatment-naïve PMR patients are allocated in a 1:1 ratio to one of two intervention arms: (1) a structured nurse-led treat-to-target prednisolone taper regimen in a hospital setting, which moving forward will be referred to as structured taper or (2) usual care prednisolone tapering in general practice (GP), which moving forward will be referred to as usual care. A detailed

description of the two intervention arms can be found in the study protocol. The study was set to include a total of 120 participants.

2.2 Randomisation

Randomisation was performed using a computer-generated randomisation schedule stratified by study site (6 levels) using permuted blocks of random sizes between 2 and 6. Block sizes were not disclosed to the investigators to ensure adequate allocation concealment.

2.3 Sample Size and Power Considerations

Assuming a 30% points difference (between proportions), an alpha of 0.05, and a statistical power of 90%, a sample size calculation based on the *guestimated* proportions of patients successfully withdrawn from prednisolone and also in remission after one year (80% vs 50%), yields 52 patients in each group. We aimed for inclusion and randomisation (1:1) of 120 participants in the intention-to-treat population (approximately 60 patients in each group).

2.4 Framework

A trial with the primary objective of showing that the response to the experimental intervention is superior to the comparative agent (control group). The hypothesis testing framework of this trial is to test if a structured prednisolone taper in a hospital setting is superior to usual care tapering in general practice for reaching prednisolone-free remission after one year.

2.5 Statistical Interim Analysis and Stopping Guidance

No statistical interim analysis is planned, and there is no guidance for stopping the trial.

2.6 Timing of Final Analysis

All primary, key secondary, and other secondary outcomes described in this SAP will be analysed collectively and reported in the primary manuscript. The final analysis will take place when all patients have completed the one-year follow-up visit (expected in October 2025).

2.7 Timing of Outcome Assessment

The timing of visits is described in Table 3 in the study protocol. The inclusion visit (Week 0) is set as the start of the study. Randomisation must take place within 2 weeks after inclusion. Questionnaires and biobank blood samples are scheduled every 3 months, corresponding to Week 13, Week 26, Week 39 and Week 52 and are aimed to be completed within 14 days. The one-year visit is set for 52 weeks after inclusion and is aimed to be completed within 14 days.

3. Statistical Principles

3.1 Confidence Intervals and P-values

All 95% confidence intervals (95% CIs) and *P*-values will be two-sided. Superiority is defined as $p<0.05$ for the primary endpoint. We will not apply explicit adjustments for multiplicity, rather we will analyse all key secondary outcomes and then interpret the key secondary outcomes in a prioritized order (i.e., "gatekeeping procedure"): The interpretation of the key secondary outcomes will be performed and interpreted in sequence until one of the analyses fails to show the statistically significant difference, or until all analyses have been completed at a statistical significance level of 0.05. The order in which key secondary outcomes will be interpreted is as listed in this SAP. Among the secondary endpoints, the 95% confidence intervals will not be adjusted for multiplicity and should not be used in place of hypothesis testing.

3.2 Intercurrent Events

Intercurrent events are defined as events that occur after randomisation but before the primary outcome measurement, which may affect the interpretation of the treatment effect. For this study, the following list of intercurrent events is defined, applicable for both intervention groups:

- GCA diagnosis before primary outcome assessment at week 52
- Change of diagnosis to a diagnosis other than PMR before primary outcome assessment at week 52
- Death before primary outcome assessment at week 52
- Withdrawal from the study before primary outcome assessment at week 52

Because of the substantial difference between the two interventions described in the protocol, certain intercurrent events with a possible direct bearing on the primary outcome are only applicable to one of the intervention groups. These are described separately: (1) For the patients randomised to usual care, the following specific intercurrent events are defined:

- Patient NOT being discharged from the hospital within 14 days of randomisation and treatment initiation
- Patient being re-referred to the department of rheumatology directly after discharge (Patients re-referred later during the tapering period due to initial treatment-resistant disease, continuous relapses, suspicion of giant cell arteritis, or other need of rheumatological assistance will not be considered intercurrent events as this is part of normal clinical practice.)

(2) For the patients randomised to structured taper, the following specific intercurrent events are defined:

- Patients discharged from the hospital out-patient clinic during the first year after randomisation and treatment initiation without having been in prednisolone-free remission
- Limited engagement with scheduled follow-ups, defined as:
    - Fewer than four phone consultations within the first year

AND/OR

- Fewer than eight blood sample assessments within the first year

The number (and percentage) of patients with intercurrent events will be summarised by treatment group. The patients included in the ITT analysis data set will be used as the denominator to calculate the percentages. No formal statistical testing will be undertaken.

3.3 Analysis Population

Primary estimand: The main analyses will be conducted in the intention-to-treat (ITT) population, emphasizing the treatment policy estimand to quantify the average treatment effect among all randomly assigned patients, regardless of treatment adherence or initiation of rescue interventions. The ITT population will include all randomised patients, regardless of their adherence to the protocol, according to the intervention they were randomised to receive.

Hypothetical estimand: To supplement the main analysis, a hypothetical estimand will be defined to assess the treatment effect among patients who adhered to the assigned intervention. This estimand is based on the principal stratum strategy and includes only patients who, up to specific time points, have not experienced any of the predefined intercurrent events. This estimand reflects the effect of treatment under optimal adherence conditions but may be subject to confounding bias (i.e., since the original randomisation is not completely respected).

4. Trial Population

4.1 Screening Data

This trial includes both a pre-screening and a screening with the following definitions: (1) Pre-screening is defined as a referred individual suspected of PMR assessed for eligibility by the investigating rheumatologist before written consent and data collection. Screening is defined as an individual suspected of PMR who is deemed eligible at pre-screening, has signed the consent form

and hereafter a full assessment of baseline data is collected to ensure eligibility. The screening period will be presented and defined from first patient, first visit to last patient, first visit. Presentation of screening data will comprise screening period, number of patients pre-screened, number of patients ineligible at pre-screening, reasons for pre-screen ineligibility, number of patients screened, number of patients ineligible at screening, reasons for screen ineligibility, and number of patients included.

4.2 Eligibility

The following inclusion criteria are applied:

- Patients newly diagnosed with PMR according to the EULAR criteria for PMR.
- No sign of GCA on ultrasonography of the temporal and axillary arteries.
- Age over 50 years.
- Danish spoken and written language skills sufficient to fill out questionnaires.

The following exclusion criteria are applied:

- Peroral, intraarticular or intramuscular application of glucocorticoids within the last month.
- Previous prednisolone treatment for GCA/PMR.
- Unable to give consent.
- Symptoms of GCA (newly onset-headache, tenderness of the temporal artery, jaw claudication, vision disturbances).
- Active malignant cancers within the last 5 years (except basal cell carcinoma).
- Other inflammatory rheumatic diseases (eg. rheumatoid arthritis, polymyositis, spondyloarthritis, psoriatic arthritits, gout).
- Uncontrolled diseases (eg severe active asthma, cardiac disease with NYHA class IV).

4.3 Recruitment

A CONSORT participant flow diagram will be drawn (Figure 1) following the CONSORT standards and will include the number of patients who were:

- Pre-screened for eligibility
- Ineligible at pre-screening*
- Eligible at pre-screening and screened
- Ineligible at screening*
- Eligible at screening and randomised
- Allocated to each intervention
- Lost to follow-up at 52 weeks*
- Randomised and included in the primary analysis (ITT population)

*reasons will be provided.

4.4 Withdrawal/Follow-up – Level of Withdrawal

Withdrawal in this study is defined solely as loss to follow-up due to withdrawal of consent, as participants are not discontinuing an active treatment but rather ceasing further participation in their assigned tapering strategy or data collection. Participants who withdraw consent will no longer contribute follow-up data beyond the withdrawal date; however, previously collected data will remain included in the analysis in accordance with the intention-to-treat (ITT) principle.

The timing of withdrawal will be recorded and summarized descriptively. Withdrawal status will be incorporated into the CONSORT flow diagram, with details on the number and percentage of participants lost to follow-up before the primary outcome assessment at week 52.

Where available, reasons for withdrawal will be documented and summarized descriptively by treatment group.

All randomised participants will be included in the main (ITT) analyses. Missing data due to withdrawal will be handled as specified in the missing data handling section.

4.5 Baseline Patient Characteristics

Baseline patient characteristics for each intervention group will be presented with the variables outlined in Table 1. Categorical data will be presented by numbers and percentages. Continuous data will be presented by mean and standard deviation (SD) if data are normally distributed and median and inter quartile range (IQR) if data are skewed. Tests of statistical significance will not be undertaken for baseline characteristics; rather the clinical importance of any imbalance will be noted.

5. Analysis

5.1 Outcome Definitions

Primary outcome measure: Prednisolone-free remission is measured at the 52-week visit and is defined as prednisolone at 0mg/day on the day of the visit combined with remission of PMR. Remission of PMR is defined as clinical remission of PMR judged from symptoms and clinical examination by the investigator at the 52-week visit combined with a CRP level below 8 mg/L. CRP levels higher than 8mg/L are accepted if explained by other causes than PMR activity, e.g. infection.

Key secondary outcome measures:

- Patient reported prednisolone dose is reported by questionnaire in mg/day at weeks 13, 26 and 39, and at the Baseline and 52-week visit, with the dose at the 52-week visit being of primary interest

- Number of relapses during the first 52 weeks is assessed at the 52-week visit by report of the patient. A relapse is defined as an increase in symptoms and/or CRP, which has led to an increase in the daily prednisolone dose. An initial increase in prednisolone dose from 15mg to a higher dose is not defined as a relapse, as it will be interpreted as a lack of initial treatment response.

- Number of GCA patients diagnosed within the first 52 weeks is assessed at the 52-week visit by assessment of the electronic patient journal. This number does not include the patients diagnosed with GCA at the 52-week visit by vascular ultrasound

- Global VAS is assessed by questionnaire at baseline and week 13, 26, 39, and 52, with the change from baseline to the 52-week visit being of primary interest. Global VAS is assessed on a visual analogue scale of 0-100 with the question "If you were to describe your overall pain in the past 24 hours, how have you been feeling?"

Other secondary outcome measures:

- Change in PMR-AS is defined as the change from baseline to the 52-week visit. PMR-AS is calculated as defined by Leeb and Bird in their original design of the score (6).

- Number of patients with an undiagnosed GCA at the 52-week visit is defined as patients who have not received a GCA diagnosis during the first 52 weeks, but who have a vascular ultrasound scan at the 52-week visit showing vascular changes diagnostic for GCA.

- SF-36 MCS is assessed by questionnaire at baseline, week 13, 26, 39 and 52 with the change from baseline to week 52 being of primary interest.

- SF-36 PCS is assessed by questionnaire at baseline, week 13, 26, 39 and 52 with the change from baseline to week 52 being of primary interest.

- HAQ-DI is assessed by questionnaire at baseline, week 13, 26, 39 and 52 with the change from baseline to week 52 being of primary interest.

- PMR VAS is assessed by questionnaire at baseline, week 13, 26, 39 and 52 with the change from baseline to week 52 being of primary interest. PMR-VAS is assessed on a visual analogue scale of 0-100 with the question "How much has your PMR affected you in the last 24 hours?"

- Fatigue VAS is assessed by questionnaire at baseline, week 13, 26, 39 and 52 with the change from baseline to week 52 being of primary interest. Fatigue VAS is assessed on a

visual analogue scale of 0-100 with the question "How much has fatigue affected you in the last 24 hours?"

- Stiffness VAS is assessed by questionnaire at baseline, week 13, 26, 39 and 52 with the change from baseline to week 52 being of primary interest. Stiffness VAS is assessed on a visual analogue scale of 0-100 with the question "How intense is your morning stiffness?"
- Morning stiffness duration is reported in minutes and is assessed by questionnaire at baseline, week 13, 26, 39 and 52 with the change from baseline to week 52 being of primary interest.
- Number of patients with DEXA scans performed within the first 3 months is assessed at the 52-week visit by assessment of the electronic patient journal.
- Number of patients with HgbA1C blood samples performed within the first 52 weeks is assessed at the 52-week visit by assessment of the electronic patient journal.
- Number of patient-reported infections is assessed by questionnaire with the question "Have you received treatment with antibiotics or medications for fungal or viral infections in the past three months?" at weeks 13, 26, 39, and 52, with the accumulated number of infections during 52 weeks being of primary interest.

5.2 Analysis Methods

For both the primary, key secondary, and other secondary outcomes, the main analysis will be based on data from the ITT population, which includes all randomised patients. Efficacy endpoints will be analysed using the full analysis set, including all randomly assigned participants with baseline measurements available, regardless of treatment initiation.

Categorical efficacy endpoints at 52 weeks will be assessed using logistic regression, with randomised treatment and clinical site as factors. To derive an adjusted risk difference with 95%CI from the odds ratio with a 95% CI, the baseline risk in the control group will be estimated

from the observed event rate in the usual care group. The risk difference will be calculated as the absolute change in probability between the structured taper and usual care groups, with confidence intervals obtained by applying the same transformation to the odds ratio's confidence limits. This method provides an interpretable measure of absolute risk while maintaining statistical adjustments from the logistic regression model. Although the trial is designed as a superiority trial, if the null hypothesis cannot be rejected, we will explore whether the results suggest equivalence in efficacy between structured taper and usual care at week 52, based on the primary endpoint. Interpretation of potential null findings (7): A possible equivalence would be considered if the 95% CI for the treatment difference falls within ±20% points at week 52. That is, for the primary endpoint, a 95% CI excluding differences greater than 20 percentage points between groups will be interpreted as indicating no clinically meaningful difference.

Continuous outcomes will be analysed as change from baseline using repeated-measures mixed-effects linear models, with patient identification as a random effect. Fixed effects will include baseline score, treatment group (structured taper or usual care), time (baseline, 13, 26, 39, and 52 weeks), clinical site (Aarhus, Gødstrup, Silkeborg, Horsens, Aalborg or Hjørring), and treatment-by-time interactions. Average changes from baseline will be reported as least-squares means with standard errors, and between-group differences will be presented as least-squares mean differences with 95% CIs. To compare the number of relapses per patient over 52 weeks, the main analysis will use negative binomial regression, which appropriately accounts for count data and overdispersion (i.e., variance exceeding the mean). A nonparametric comparison will be considered as a supportive or exploratory analysis if model assumptions appear violated. The primary effect estimate will be the rate ratio (with 95% CI) comparing structured taper with usual care.

5.3 Missing Data

For categorical efficacy endpoints, missing data will be handled simplistically using non-responder imputation (NRI), a conservative approach commonly applied in randomised trials evaluating treatment response rates. NRI assumes that participants with missing categorical outcome data are

non-responders, thereby minimizing the risk of overestimating treatment efficacy. This method is particularly appropriate in regulatory settings, such as trials of new therapies/strategies, where rigorous control of potential bias is essential. The sensitivity of the estimates derived from the logistic regressions using simplistic non-responder imputation, will be assessed by conducting a tipping point analysis. The analysis will be carried out using best-case, worst-case, best-worst, and worst-best imputations. In this analysis a range of assumptions are explored and seeks to identify the conditions under which the primary conclusion will change. Thus, the robustness of the results is tested and the degree of dependence on missing data assumptions is quantified.

For continuous efficacy outcomes, missing data will be handled with the use of the mixed-effects linear-models approach with an assumption that data is missing at random (MAR), in accordance with the working assumption underlying these models (9). Unlike traditional simplistic imputation methods, mixed-effects models use all available data points to estimate treatment effects while accounting for within-subject correlation and variability over time. This approach reduces bias and maintains statistical power by leveraging maximum likelihood estimation, which provides unbiased parameter estimates under MAR. Additionally, mixed-effects models appropriately handle unequal time intervals and accommodate variability across participants by incorporating random effects, making them particularly well-suited for longitudinal clinical trial data. To evaluate the robustness of the estimates from the mixed-effect linear model, a sensitivity analysis will be performed. The missing values will be imputed using baseline observation carried forward and the same mixed-effect linear model will be used. This approach will help assess the influence of the MAR assumption on the observed treatment effects.

5.4 Harms

Data on adverse events (AE) will be collected on two levels in this study. First, patients will complete a questionnaire every three months during the study period, including known possible side effects of prednisolone including patient-reported infections requiring antibiotics, antiviral or antifungal treatment. If a specific AE is reported more than once from the same patient during the study period, it will only be counted once in the final analysis. Second, serious adverse events (SAE)

as defined by the International Council for Harmonisation – Good Clinical Practice (ICH-GCP) are recorded at the 52-week visit, including an assessment by the investigator of whether the SAE is related to prednisolone. The number (and percentage) of patients experiencing each AEs/SAEs, as well as the number (and rate) of events, will be presented for both treatment groups of the ITT population (Table 3). No formal statistical testing will be undertaken.

5.5 Statistical Software

The statistical analysis will be carried out using STATA version 18.5 (StataCorp LCC, College Station, TX, USA) and R version 4.4.2 (R Core Team, Vienna, Austria) using the packages lme4 and emmeans.

6. Tables and Figures

Figure 1

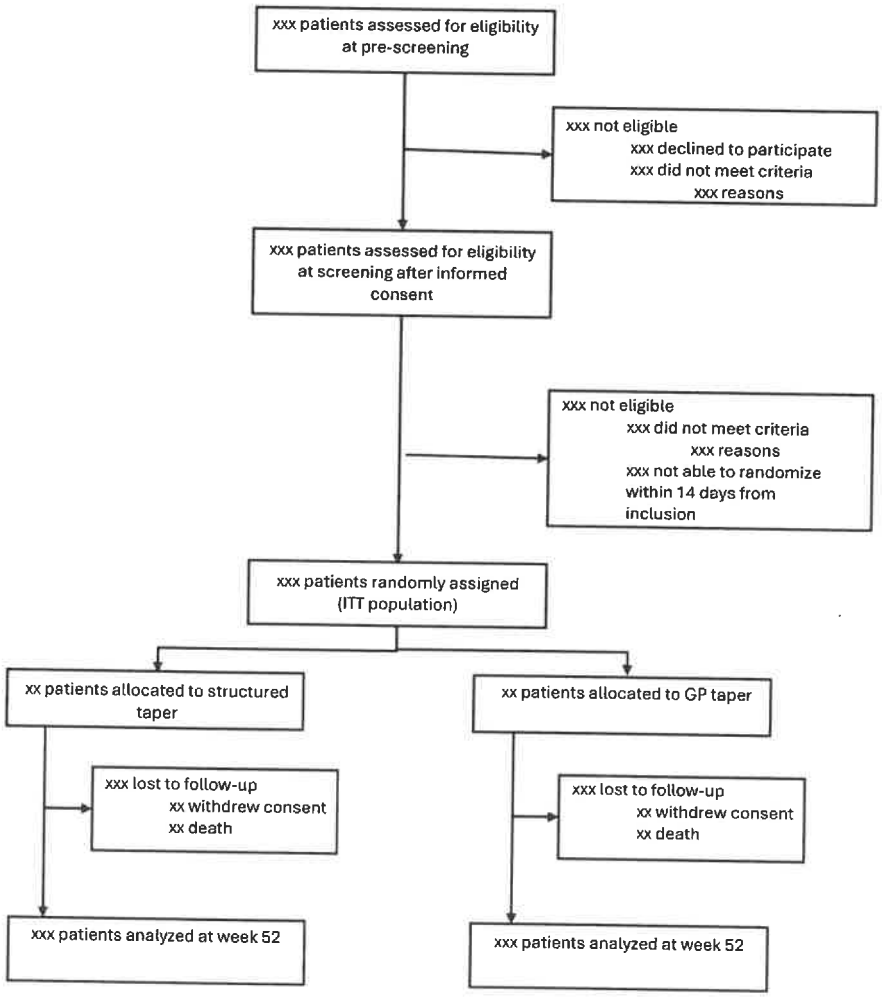

Figure 2

Figure 2 will be presented as a multi panel figure with Figure 2A being a bar graph with 95%CI showing the proportion of patients in prednisolone-free remission at the 52-week visit in each treatment group (Primary outcome) and Figure 2B being model-based estimates of the change in daily prednisolone dose (mg) from baseline to weeks 13, 26, 39, and 52, for each treatment group derived from the linear mixed-effects model . Least-squares means with vertical 95% confidence intervals are plotted for each treatment group at each time point. (1st Key Secondary outcome)

Table 1. Baseline Patient Characteristics in the ITT population

	Structured taper (n=)	Usual care (n=)	Total (n=)
Demographics			
Age (years)			
Female sex, n(%)			
Body Mass Index (kg/m^2)			
Symptoms			
Symptom duration (weeks)			
Bilateral Shoulder pain, n(%)			
Shoulder Pain VAS			
Bilateral Hip pain, n(%)			
Hip Pain VAS			
Morning stiffness >45 min, n(%)			
New onset fatigue, n(%)			
Unintended weight loss >2kg, n(%)			
Joint pain besides hip and shoulders, n(%)			
Clinical examination			
Restricted shoulder range of motion ≥1 shoulder, n(%)			
Restricted hip range of motion ≥1 hip, n(%)			
Tender joint count ≥1 , besides shoulders and hips, n (%)			
Swollen joint count ≥1 , besides shoulders and hips, n (%)			
Lab results			
CRP (mg/L)			
Normal Rheumatoid Factor and Anti-CCP, n(%)			
Outcome Measures			
Global Pain VAS			
PMR Activity Score			
SF-36 MCS			
SF-36 PCS			
HAQ-DI			
PMR VAS			
Fatigue VAS			
Stiffness VAS			
Morning stiffness duration (min)			

Values will be reported as means and standard deviations unless otherwise indicated.

Table 2 – Primary and Secondary Endpoints at Week 52

	Structured taper (n=)	Usual care (n=)	Difference Between Groups (95%CI)	p-value
Primary Endpoint				
Patients in prednisolone-free remission, n(%)				
Key Secondary Endpoints				
Daily prednisolone dose change from baseline (mg/day)				
Number of relapses per patient during 52 weeks, counts (rates)*				
GCA diagnosed before the 52-week visit, n (%)				
Global Pain VAS change from baseline				
Secondary Endpoints				
PMR-AS change from baseline				
GCA diagnosed at the 52-week visit by vascular US, n(%)				
SF-36 MCS change from baseline				
SF-36 PCS change from baseline				
HAQ-DI change from baseline				
PMR VAS change from baseline				
Fatigue VAS change from baseline				
Stiffness VAS change from baseline				
Morning stiffness duration change from baseline (min)				
Patients with DEXA scan performed within 3 months of diagnosis, n(%)				
Patients with HgBA1c performed within first 52 weeks, n(%)				

Estimates will be reported as least squares means with standard errors unless otherwise indicated. The 95% confidence intervals will not be adjusted for multiplicity and should not be used in place of hypothesis testing. *To compare the number of relapses per patient over 52 weeks, the effect estimate will be the rate ratio (with 95% CI) comparing structured taper with usual care.

Table 3 – AEs and SAEs

	Structured taper (n=)	Usual care (n=)	Risk Difference (95%CI)
Patient reported AEs			
AE, n patients (%)			
AE, n events (rate – events per participants month)			
Patient reported infections, n events (rate – events per participants month)			
Withdrawal due to adverse events, n (%)			
Serious Adverse Events			
SAE, n patients (%)			
SAE, n events (rate)			
SAEs relationship to prednisolone			
Probably not related, n events (rate)			
Probably related, n events (rate)			
Deaths, n (%)			

SUPPLEMENTARY, PRESPECIFIED ANALYSES

Table S1. Hypothetical Estimand: Primary and Key Secondary Endpoints at Week 52

	T2T Taper (n=)	GP Taper (n=)	Difference Between Groups (95%CI)
Primary Endpoint			
Patients in prednisolone-free remission, n(%)			
Key Secondary Endpoints			
Daily prednisolone dose change from baseline (mg/day)			
Number of relapses per patient during 52 weeks, counts (rates)*			
GCA diagnosed before the 52-week visit, n (%)			
Global Pain VAS change from baseline			

Estimates will be reported as least squares means with standard errors unless otherwise indicated.

* To compare the number of relapses per patient over 52 weeks, the effect estimate will be the rate ratio (with 95% CI) comparing structured taper with usual care.

Table S2. Sensitivity analysis: Primary and Key Secondary Endpoints at Week 52

	T2T Taper (n=)	GP Taper (n=)	Difference Between Groups (95%CI)
Primary Endpoint			
Patients in prednisolone-free remission, n (%)			
Key Secondary Endpoints			
Daily prednisolone dose change from baseline (mg/day)			
Number of relapses per patient during 52 weeks, counts (rates)*			
GCA diagnosed before the 52-week visit, n (%)			
Global Pain VAS change from baseline			

Missing continuous outcomes will be simplistically imputed as baseline observation carried forward. Categorical outcomes will be imputed using best-case, worst-case, best–worst, and worst–best scenarios to support a tipping point analysis. Estimates will be reported as least squares means with standard errors unless otherwise indicated. *To compare the number of relapses per patient over 52 weeks, the effect estimate will be the rate ratio (with 95% CI) comparing structured taper with usual care.

7. References

1. Buttgereit F, Dejaco C, Matteson EL, Dasgupta B. Polymyalgia Rheumatica and Giant Cell Arteritis: A Systematic Review. Jama. 2016;315(22):2442-58.
2. Dejaco C, Singh YP, Perel P, Hutchings A, Camellino D, Mackie S, et al. 2015 Recommendations for the management of polymyalgia rheumatica: a European League Against Rheumatism/American College of Rheumatology collaborative initiative. Ann Rheum Dis. 2015;74(10):1799-807.
3. Floris A, Piga M, Chessa E, Congia M, Erre GL, Angioni MM, et al. Long-term glucocorticoid treatment and high relapse rate remain unresolved issues in the real-life management of polymyalgia rheumatica: a systematic literature review and meta-analysis. Clin Rheumatol. 2022;41(1):19-31.
4. Mørk C, Dam MY, Callsen MG, Keller KK. The effect of a nurse-led prednisolone tapering regimen in polymyalgia rheumatica: a retrospective cohort study. Rheumatol Int. 2020.
5. HOCHBERG Y. A sharper Bonferroni procedure for multiple tests of significance. Biometrika. 1988;75(4):800-2.
6. Leeb BF, Bird HA. A disease activity score for polymyalgia rheumatica. Ann Rheum Dis. 2004;63(10):1279-83.
7. Christensen R, Ranstam J, Overgaard S, Wagner P. Guidelines for a structured manuscript: Statistical methods and reporting in biomedical research journals. Acta Orthop. 2023;94:243-9.
8. Liublinska V, Rubin DB. Sensitivity analysis for a partially missing binary outcome in a two-arm randomized clinical trial. Stat Med. 2014;33(24):4170-85.
9. Detry MA, Ma Y. Analyzing Repeated Measurements Using Mixed Models. Jama. 2016;315(4):407-8.