Biostatistics Research Group,
Population Health Sciences Institute, Newcastle University

Randomised controlled trial of plasma cell depletion for severe Graves' disease

---

Statistical Analysis Plan for Graves- PCD

01/07/2024

SAP Version number: 2.0

This statistical analysis plan is based on protocol version 6.0 [18/05/2023]

Sponsor: The Newcastle upon Tyne Hospitals NHS Foundation Trust

Funded by: Medical Research Council

Funder ID: MR/V005898/1

EudraCT Number: 2020-005635-78

ISRCTN: 81162400

NHS REC Reference: 21/LO/0449

IRAS ID Number: 1003652

Prepared by:

Name	Michael Cole	Role	Trial Statistician
Signature		Date	20/08/2024

Authorised by:

Name	James Wason	Role	Senior Statistician
Signature		Date	01/07/2024
Name	Simon Pearce	Role	Chief Investigator / Clinical lead
Signature		Date	01/07/2024

This current version of the SAP and all preceding versions will be stored in the Statistical Section of the Trial Master File held by the PHSI Biostatistics Research Group.

Revision history

Version	Date	Changes made	Justification for change	Timing of change
1.0	13/10/2022	First version	First version of SAP to provide details of interim analysis. Details of primary analysis will be included in subsequent versions of this SAP	
1.1	14/10/2022	Minor change to table layout and baseline characteristics, plus wording clarification in section 4.3.2 part 1.		
2.0	01/07/2024	Final version of SAP for full analysis. Major changes made to all sections corresponding to the final analysis.		

Glossary of abbreviations

ABBREVIATION	DEFINITION
AE	Adverse Event
AR	Adverse Reaction
ATPO	Antibodies to thyroid peroxidase
BMI	Body Mass Index
BTF	British Thyroid Foundation
CA	Competent Authority
CAS	Clinical Activity Score
CBZ	Carbimazole
CDMS	Clinical data management system
CI	Chief Investigator
COPD	Chronic obstructive pulmonary disease
CTA	Clinical Trial Authorisation
CTIMP	Clinical Trial of an Investigational Medicinal Product
DMC	Data Monitoring Committee
DPFS	Developmental Pathway Funding Scheme
DR	Dose-response
DSUR	Development Safety Update Report
eCRF	Electronic Case Report Form
ECG	Electrocardiogram
EDTA	Ethylenediaminetetraacetic acid
EoT	End of trial
EudraCT	European Clinical Trials Database
FACS	Fluorescence Activated Cell Sorting
FBC	Full Blood Count
FEV	Forced Expiratory Volume
FSH	Follicle stimulating hormone
FT3	Serum free tri-iodothyronine
FT4	Serum free thyroxine
GCP	Good Clinical Practice
GOQoL	Graves' Ophthalmopathy Quality of Life Questionnaire
GP	General Practitioner
HBsAG	Hepatitis B surface antigen
HCG	Human Chorionic Gonadotropin

HCV	Hepatitis C Virus
HIV	Human Immunodeficiency Virus
hr	Hour
HRA	Health Research Authority
ICH	International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human use
IRR	Infusion related reactions
Ig	Immunoglobulin
IMP	Investigational Medicinal Product
ITT	Intention to treat
IUD	Intrauterine device
IUS	Intrauterine hormone-releasing system
kg	Kilogram
LFT	Liver Function Test
LPLV	Last Patient Last Visit
mAb	Monoclonal antibody
MedDRA	Medical Dictionary for Regulatory Activities
MHRA	Medicines and Healthcare products Regulatory Agency
mg	Milligram
ml	Millilitre
mRNA	Messenger ribonucleic acids
NCTU	Newcastle Clinical Trials Unit
NHS	National Health Service
NIMP	Non-investigational medicinal product
NUTH	The Newcastle upon Tyne Hospitals NHS Foundation Trust
PBMC	Peripheral blood mononuclear cell
PI	Principal Investigator
PIC	Participant Identification Centre
PIS	Participant Information Sheet
PP	Per-protocol
PR	Percentage reduction
PTU	Propylthiouracil
QA	Quality Assurance
R&D	Research & Development
REC	Research Ethics Committee
RSI	Reference Safety Information

SAE	Serious Adverse Event
SAP	Statistical analysis plan
SAR	Serious Adverse Reaction
SARS-CoV2	Severe acute respiratory syndrome coronavirus 2
SI	Statutory Instruments
SOP	Standard Operating Procedure
SmPC	Summary of Product Characteristics
SUSAR	Suspected Unexpected Serious Adverse Reaction
TB	Tuberculosis
TBII	Thyrotropin-binding Inhibitory Immunoglobulin
TED	Thyroid Eye Disease
Tg	Thyroglobulin
TgAB	Thyroglobulin antibodies
ThyPRO	Thyroid Patient Reported Outcome
TMG	Trial Management Group
TRAb	TSH receptor antibodies
TSC	Trial Steering Committee
TSH	Thyroid Stimulating hormone
TMF	Trial Master File
U & E	Urea & Electrolytes
USM	Urgent Safety Measure
USS	Ultrasound Scan
WHO	World Health Organisation

Table of Contents

1 Introduction 8

1.1 Background and rationale 8

1.2 Trial Objectives and outcome measures 9

1.2.1 Primary objective 9
1.2.2 Secondary objectives 9
1.2.3 Exploratory objectives 9
1.2.4 Primary outcome measure 9
1.2.5 Secondary outcome measures 9
1.2.6 Exploratory objectives / outcome measures 10

2 Study Methods 11

2.1 Trial design 11

2.2 Trial setting and patient population 11

2.3 Inclusion Criteria 12

2.4 Exclusion Criteria 12

2.5 Randomisation and blinding 13

2.6 Sample size and power 13

2.7 Study Diagram/Flowchart 14

2.8 Trial timelines 15

2.9 Target accrual 15

3 Data Collection And Outcome Measures 16

3.1 Trial assessments 16

3.1.1 Primary outcome 16
3.1.2 Secondary outcomes 16

3.2 Definition and calculation of outcome measures 17

3.2.1 Primary outcome 17
3.2.2 Secondary outcomes 18
3.2.3 Dose of IMP received (mg/kg) 23
3.2.4 Dose of NIMP received 24

4 Trial Population 25

4.1 Patient flow through trial 25

4.2 Example CONSORT flow diagram 26

4.3 Recruitment 27

4.4 Analysis populations 27

4.5 Baseline characteristics 28

4.6 Treatment received 29

4.6.1 Daratumumab 29
4.6.2 ATD 29
4.6.3 Thyroid concomitant medications 30
4.6.4 Non-thyroid concomitant medications 31

4.7 Protocol deviations 32
5 Interim Analyses, Data Monitoring And Stopping Guidelines 33
5.1 Dose-response modelling 33
5.2 Model-based interim analysis dosing decisions 34
5.3 Dosing decisions, other considerations 35
5.4 Contents of interim analysis report 35
5.4.1 Dose response modelling results 35
5.4.2 Individual participant details 36
6 Statistical considerations 38
6.1 Timing of analyses 38
6.2 Analysis Methods 38
6.2.1 Analysis of primary outcome 38
6.2.2 Secondary analysis of primary outcome 40
6.2.3 Non-linear dose-response models 41
6.2.4 Subgroup Analyses 42
6.2.5 Analysis of secondary outcomes 43
6.2.6 Additional analyses 47
7 Safety 48
7.1 Adverse reactions 49
7.2 Serious adverse events 52
7.3 Other safety data 52
8 Statistical Software 54
9 References 54
10 Appendix 55
10.1 Indicative normal ranges pertinent to Stage 1 interim analysis 55
10.2 Scoring of the ThyPRO-39 questionnaire 56

1 Introduction

1.1 Background and rationale

Graves' disease (autoimmune hyperthyroidism) affects around 3% of women and 0·5% of men over a lifetime, and most commonly presents in the 4th and 5th decades of life, with a disproportionate burden of ill-health falling on working-age women. Typical symptoms include weight loss, palpitations, breathlessness, sweating, heat intolerance, tremor, insomnia, loss of concentration and irritability. Around 40% of patients develop an inflammatory eye condition, Thyroid Eye Disease (TED) which can cause facial disfigurement as well as functional visual problems and loss of sight. Rarely, a specific skin complication, Thyroid Dermopathy (aka Pretibial Myxoedema) may occur leading to a brawny thickening of the skin on the lower legs and feet. These latter 2 problems produce particularly distressing symptoms for a relatively young and active patient group.

The usual treatments for the hyperthyroidism of Graves' disease are either antithyroid drugs (carbimazole - CBZ), radio-active iodine or surgical thyroidectomy. Antithyroid drugs lead to remission in only around 50% of people. Around 10% of patients have severe Graves' disease, defined by severe thyrotoxicosis (serum FT4 ≥50pmol/l or FT3≥15pmol), failure of medical control of hyperthyroidism, large goitre, inflammatory thyroid eye disease or thyroid dermopathy. They are characterised by high concentrations of the directly pathogenic TSH-receptor stimulating antibodies (TRAb), and 80% of them relapse following conventional medical therapy with antithyroid drugs. Quality of life in thyroid eye disease patients is poor, worse than for diabetes, and similar to patients with inflammatory bowel disease. Current treatments for these patients give unsatisfactory outcomes and are expensive, typically involving multiple episodes of eye surgery and/or surgical thyroidectomy. This study will determine proof of concept that the plasma cell depleting antibody daratumumab can ameliorate severe Graves' disease, using TRAb concentrations and circulating thyroid hormone levels along with clinical disease severity/activity scores as outcome measures.

Severe Graves' disease is caused by high titres of directly pathogenic TRAb, which are secreted from terminally differentiated B lymphocytes known as plasma cells. Failure of medical control and/or early relapse following conventional antithyroid drugs reflects persistence of these long-lived, TRAb-secreting plasma cells in the secondary lymphoid tissues and bone marrow. Both benign and malignant plasma cells express high levels of the cell-surface glycoprotein CD38. Daratumumab, a monoclonal antibody that binds to CD38 was recently licenced for the plasma cell malignancy, myeloma, has the potential to deplete plasma cells and produce a rapid reduction in TRAb levels, which may alter the natural history of severe Graves' disease. This trial aims to establish proof of concept that daratumumab has efficacy in severe Graves' disease patients and will provide important data to inform a choice of dosing regimen.

Daratumumab is licenced for use in myeloma at a dose of 16mg/kg, but the optimal dose in patients with Graves' disease is currently unknown. In contrast to patients with myeloma, those with Graves' disease have several orders of magnitude fewer plasma cells and it is therefore expected that lower doses of drug will be active in this patient group with benign disease. In addition, patients with Graves' disease will be younger (median age of presentation is 40 years) and in better overall health (bearing in mind study exclusion criteria) than patients with myeloma who had to have refractory disease to several previous treatments to enter the early-phase trials of daratumumab. Dose-finding trials in myeloma patients showed no difference in adverse event rates between doses of 8mg/kg and 16mg/kg, and no dose-limiting toxicity up to

24mg/kg. Therefore, in order to define a signal for efficacy this study will use 9mg/kg as the top dose in stage 1, with reducing concentrations (3mg/kg, 1mg/kg and 0.5mg/kg) along with placebo to determine the dose response in stage 1.

1.2 Trial Objectives and outcome measures

1.2.1 Primary objective

- To determine if daratumumab modulates the humoral immune response in Graves' disease patients

1.2.2 Secondary objectives

- To determine how fast daratumumab modulates the humoral immune response in Graves' disease patients
- To determine the optimal dose (or dose range) of daratumumab for Graves' disease patients
- To determine if daratumumab reduces thyroid hormone levels
- To determine if daratumumab changes the time course of serum TSH
- To determine if daratumumab changes thyroid size
- To determine if daratumumab changes other thyroid autoantibodies
- To determine if daratumumab improves thyroid eye disease
- To determine if daratumumab improves thyroid symptom related QoL
- To determine if daratumumab is safe in this patient group

1.2.3 Exploratory objectives

- To determine if daratumumab changes lymphocyte/plasma cell transcriptomic markers
- To determine if daratumumab changes the lymphocyte subsets

1.2.4 Primary outcome measure

- Change in serum TRAb antibodies from baseline to 12 weeks compared to change in placebo group

1.2.5 Secondary outcome measures

- Change in serum TRAb antibodies from baseline to 2, 4, 6, 12 and 24 weeks
- Dose-response curve for daratumumab against change in serum TRAb antibodies from baseline to 6 and 12 weeks
- Change in serum FT3 and FT4 from baseline to 2, 4, 6, 12 and 24 weeks
- Change in serum TSH from baseline to 2, 4, 6, 12 and 24 weeks
- Change in thyroid volume from baseline to 24 weeks measured by ultrasound
- Change in serum ATPO and thyroglobulin antibodies from baseline to 6, 12 and 24 weeks

- Change in CAS, composite eye index and GOQoL score from baseline to 6, 12 and 24 weeks
- Change in ThyPRO39 score from baseline to 6, 12 and 24 weeks
- Change in serum immunoglobulins, specific antibodies including (SARS-CoV2) and blood count parameters from baseline to 6, 12 and 24 weeks.
- Adverse Reactions to 24 weeks

1.2.6 Exploratory objectives / outcome measures

- Analysis of blood plasma cell markers and mRNA signature
- Change in lymphocyte subsets (by FACS) from baseline to 6, 12 and 24 weeks

2 Study Methods

2.1 Trial design

Graves-PCD is an adaptive, 2-stage randomised phase IIa clinical trial that will recruit 30 patients with severe Graves' disease from NHS secondary care. It is a single blinded trial in which participants will be blind to allocation.

Stage 1 is a dose-response study using 4 doses of daratumumab (9mg/kg, 3mg/kg, 1mg/kg, 0.5mg/kg) and a colourless, volume-matched placebo infusion in approximately 15 patients (i.e. five groups of n=3, randomised in a 1:1:1:1:1 ratio).

Following Stage 1, an interim analysis will be performed in order to select an optimal dose(s) of daratumumab for Stage 2. ***The dose selection will be based on an analysis of the reduction in TRAb antibody concentration and safety assessed at 12 weeks.***

In Stage 2, the remaining patients will be randomised between placebo and one or two chosen doses of daratumumab depending on results of the interim analysis.

The target recruitment rate is 2 participants per month. Recruitment will continue using Stage 1 dose allocations during the 12 weeks follow up of Stage 1 participants and during the interim analysis. Recruitment to Stage 2 will then take place over a further 7 months. Patients in both stages of the trial will be followed for 24 weeks, with the primary endpoint being measured at week 12.

2.2 Trial setting and patient population

This is a single centre trial which will be undertaken at the Newcastle upon Tyne Hospitals NHS Foundation Trust (NUTH) site with the option to add another site in stage 2. Patients will be recruited during attendance at NHS specialist thyroid/ TED clinics. Participants will also be identified through local neighbouring trusts acting as participant identification centres (PICs) and by self-referral through adverts. If initial recruitment to the trial is lower than anticipated, consideration will be made to the addition of a further research site.

If after IMP dosing, avoidable individual travel is not recommended due to local or national restrictions or if the travel burden for self-referring participants would be too great as judged by the research team, then visits 6 to 8 may be performed by phone, followed by a home visit or a visit to the local health services for collecting a blood sample to ascertain safety and outcome measures. In these circumstances, treatment of newly recruited participants will also be temporarily halted, but treatment of participants who have already received the first dose of daratumumab/placebo would continue.

2.3 Inclusion Criteria

1. Patients ≥18yrs old
2. Recent-onset Graves' disease (within 12 months) (defined as date of first thyroid function test showing hyperthyroidism (FT4 and TSH) in current episode)
3. TRAb antibody concentrations above 10U/L (on Roche or Brahms TBII assays)
4. One or more of:
    - Pre-treatment severe hyperthyroidism (FT4 ≥50 pmol/L; or FT3 ≥15 pmol/l)
    - Persisting hyperthyroidism despite more than 12 weeks of antithyroid drug therapy (defined as FT3 above the upper limit of the reference range following 12 weeks of carbimazole treatment at a dose of 40mg or more daily (or equivalent dose of PTU))
    - Inflammatory thyroid eye disease (defined as clinical activity score, CAS≥3), or thyroid dermopathy
    - Large (visible) goitre (WHO grade III)
    - 2 or more relapses *(3 episodes in total)* despite completing 12 months or more of medical treatment on each occasion. Relapse is defined as FT3 above the upper limit of the local reference range.
5. For women of child-bearing potential, willing to use a highly effective contraceptive method during their participation in the trial
6. Able to understand and speak sufficient English to complete trial procedures
7. Willing and able to provide informed consent prior to any trial procedures taking place

2.4 Exclusion Criteria

1. Previous thyroidectomy, or radioiodine treatment within 2 years
2. Pregnant or breastfeeding, or with a plan for pregnancy within 6 months
3. Previous shingles, known untreated cervical dysplasia, hepatitis B & C or HIV infection
4. Anaemia (Hb ≤ 100g/l), thrombocytopenia (≤75 $x10^9$/L) or neutropenia (≤1.0 $x10^9$/L)
5. Known chronic obstructive pulmonary disease (COPD) (defined as a forced expiratory volume [FEV] in 1 second <60% of predicted normal), persistent asthma, or a history of asthma within the last 2 years (intermittent asthma without hospitalisation is allowed)
6. Any significant physical or mental health condition that impacts the safety of the intervention, the interpretation of thyroid function or the ability of a participant to attend for intervention and safety monitoring, e.g. major cardiorespiratory disease, renal or hepatic failure, pancreatitis, cancer undergoing active treatment (excluding non-

melanoma skin cancer), untreated chronic infection including TB, psychosis, depression impairing Activities of Daily Living

7. Current use of immunosuppressive therapy for thyroid eye disease or other conditions (within 3 months)
8. Current or previous participation in a CTIMP research study within 4 months
9. Hypersensitivity or anaphylactic reaction to previous monoclonal antibody treatments or methylprednisolone
10. Inability, in the opinion of the investigator, to be able to complete the clinical trial visits or procedures.

2.5 Randomisation and blinding

Stage 1: Trial participants will be randomly allocated to receive one of four different doses of daratumumab (9, 3, 1, 0.5mg/Kg) or matched placebo. Participants will be randomised in a ratio of 1:1:1:1:1 to the 4 daratumumab regimens or placebo intervention.

Stage 2: Participants will be randomly allocated to receive daratumumab (the one or two doses available at Stage 2 as determined by the interim analysis), or matched placebo. Participants will be randomised in a ratio of 1:1:1 to the 2 chosen daratumumab regimens or placebo intervention. If the interim analysis recommends only one daratumumab regimen, then patients will be randomised 1:1 between daratumumab and placebo.

Allocation sequences for both Stages will be computer-generated, using a random permuted block design; blocks might vary in size. Block sizes will not be disclosed, to ensure concealment. Randomisation will be performed by delegated and trained members of the research team using the Sealed Envelope system and should take place as close as practically possible to the baseline visit but with sufficient time to allow for prescription and medication ordering and dispensing. The allocation will be documented in the participant medical records and will be the same for each participant for both of their trial treatment visits.

The trial is single-blinded. **Participants will be unaware of their allocated treatment group.** The CI/PI and delegated clinicians as well as the TMG will be unblinded to the treatment allocation.

2.6 Sample size and power

The trial has been powered on change in TRAb antibody concentration. Daratumumab can selectively deplete antibody-secreting plasma cells to undetectable circulating numbers and so, **a 50% or larger reduction in TRAb antibody concentration within 12 weeks of daratumumab administration is biologically plausible and would be a clinically important effect**.

Data from an observational study in milder Graves' disease showed a mean change in TRAb concentration of 2% over 6 weeks, with an SD of 15%. Here, we assume a larger SD of 25% (to account for the outcome in this trial being measured at 12 weeks as opposed to 6 weeks).

The logarithm of the percentage reduction in TRAb is assumed normally distributed with SD 0.5 (based on untransformed values of 25% SD and 50% mean, i.e. CV is 0.5).

Simulations are performed using 4 different dose-response relationships:

1. plateau effect corresponds to 80% mean reduction in TRAb, 3mg/kg ED50
2. plateau effect corresponds to 60% reduction, 1mg/kg ED50
3. plateau effect corresponds to 60% reduction, 3mg/kg ED50
4. 5% mean reduction regardless of dose - null scenario.

All simulations assume an Emax dose-response and a 5% mean reduction in placebo treated participants. The ED50 is the dose that gives 50% of the difference in effect between placebo and plateau.

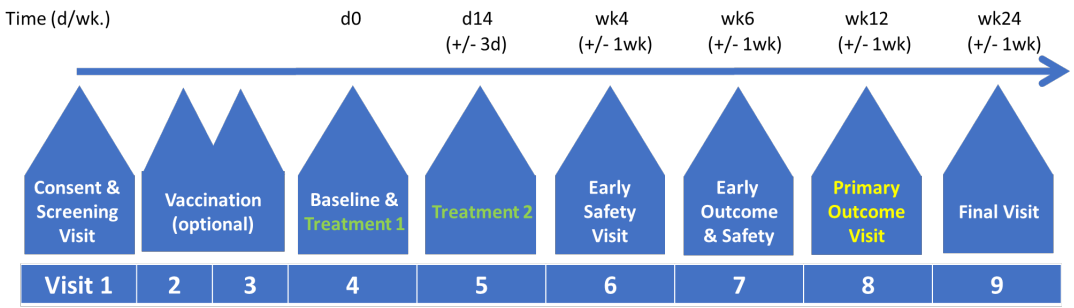
The power/type I error rate of the trial for each scenario is in the table below:

Scenario	**Probability to conclude dose-response (10000 replicates)**
1 – plateau of 80%, 3mg/kg ED50	>99%
2 – plateau of 60%, 1 mg/kg ED50	92%
3 – plateau of 60%, 3mg/kg ED50	83%
4 – null scenario	3.2%

For Scenario 2, a trial with 27 participants will have more than 90% power (one-sided a=0.05) to conclude there is a dose response relationship. In order to allow for an assumed 10% drop out, we will plan to recruit 30 participants.

Details of the interim analysis and criteria for premature termination of the trial are given in section 5.0. Simulations used in the power calculations follow this analysis plan with one exception; if the estimated plateau effect (the maximum effect possible as the dose becomes very large) lies between a 25%-50% reduction, then we have assumed, conservatively, that the trial will continue with doses of 3mg/kg, 9mg/kg plus placebo rather than doses of 9mg/kg, 12 or 16mg/Kg, and placebo as specified in section 5.3.3.

2.7 Study Diagram/Flowchart



2.8 Trial timelines

Original Duration of funding / grant	Duration: 36 months. From: 01/02/2021 to 30/01/2024
Revised end date	31/10/2024
Date of ethical favourable opinion	8/7/2021
Date of MHRA approval	21/7/2021
Date of HRA approval	21/7/2021
Date first site open to recruitment	29/9/2021
Date first participant randomised	14/10/2021
Original planned end of recruitment	31/01/2023
Revised end of recruitment	31/10/20243
Last participant recruited	28/09/2023
Planned time of formal interim analysis	November 2022
Planned time of primary analysis	June 2024 to September 2024

2.9 Target accrual

The target accrual for the trial is 30 participants.

3 Data Collection And Outcome Measures

3.1 Trial assessments

Refer to protocol section ***7. Trial Procedures*** for full details of all trial assessments.

3.1.1 Primary outcome

Change in serum TRAb antibodies from baseline to 12 weeks compared to change in placebo group. Blood samples for the primary outcome measure are taken at Visit 4 (Baseline) and Visit 8 (Primary outcome visit). Extra TRAb samples are taken and stored until data lock at these visits.

If the baseline sample is not available, TRAb at the screening visit can be used for the primary outcome provided ***the visit occurs within 28 days of the baseline visit***.

3.1.2 Secondary outcomes

Important assessments related to secondary outcomes are:

- Serum TRAb antibodies assessed at baseline and at weeks 2, 4, 6, 12 and 24.
- Serum TSH, FT3 and FT4 assessed at baseline and at weeks 2, 4, 6, 12 and 24.
- Thyroid volume assessed at baseline and 24 weeks.
- Serum ATPO and thyroglobulin antibodies assessed at baseline and at weeks 6, 12 and 24.
- Clinical Activity Score (CAS), composite eye index and GOQoL score assessed at baseline and weeks 6, 12 and 24.
- ThyPRO39 score assessed at baseline and weeks 6, 12 and 24 weeks.
- Serum immunoglobulins, specific antibodies including (SARS-CoV2) and blood count parameters at baseline and weeks 6, 12 and 24.
- Adverse Reactions to 24 weeks

3.2 Definition and calculation of outcome measures

3.2.1 Primary outcome

The primary outcome measure is an individual's change in serum TRAb antibody level from baseline to 12 weeks. TRAb antibody levels will be high at baseline (inclusion criteria requires TRAb concentrations above 10U/L) thus change will tend to be a reduction in TRAb levels.

Define $\mathrm{TRAb}^t = \mathrm{TRAb\ concentration\ at\ week\ t}$. The percentage reduction (PR) in TRAb at week *t* is given by:

$$PR_t = 100 * \frac{[\mathrm{TRAb}^0 - \mathrm{TRAb}^t]}{\mathrm{TRAb}^0}$$

The primary outcome is therefore PR_{12}

where,

TRAb^0 = TRAb at *Visit 4, day 0*
TRAb^{12} = TRAb at *Visit 8, week 12*

If TRAb^0 is unavailable, TRAb at the screening visit can be used for the primary outcome ***provided the visit occurs within 28 days of the baseline visit***.

It is feasible that TRAb concentrations at week 12, and at other times points, might be below the limit of quantification of the assay. If this situation arises, such values will be replaced by $LOQ/\sqrt{2}$ where LOQ is the limit of quantification of the assay. The use of $\mathrm{LOQ}/\sqrt{2}$ is based on the assumption that data below the LOQ follows a triangular distribution with median $\mathrm{LOQ}/\sqrt{2}$ (i.e. values drop linearly to zero density at zero concentration).

3.2.2 Secondary outcomes

3.2.2.1 Serum TRAb antibodies

The change in serum TRAb antibodies from baseline to 2, 4, 6, 12 and 24 weeks is defined in the same way as the primary outcome, i.e. percentage reduction in TRAb from baseline.

3.2.2.2 TSH, FT3 and FT4

To determine if daratumumab reduces thyroid hormone levels
To determine if daratumumab changes the time course of serum TSH
Change in serum TSH, FT3 & FT4 from baseline to 2, 4, 6, 12 and 24 weeks

Serum TSH (mU/L), FT3 (pmol/L) and FT4 (pmol/L) are assessed at baseline and at weeks 2, 4, 6, 12 and 24. TSH might be recorded as < lower limit of detection (LLD). FT3/4 might be recorded as > upper limit of detection (UL).

Upper limit of normal (ULN) for FT4 is defined as 22pmol/L, except if the participant is taking levothyroxine in which case ULN is defined as 25pmol/L.

The following outcome measures will be used for FT3 and FT4:

1. Normal level at week 12, defined as level < UL (Y/N)?
2. Normal level at week 24, defined as level < UL (Y/N)?
3. Number of weeks to normalise, defined as week number for which level first < UL (Categorical: 2, 4, 6, 12, 24).
4. Number of weeks for which level < UL - a LOCF approach will be used to calculate number of weeks.

The following outcome measures will be used for TSH:

1. Non-supressed level at week 12, defined as level ≥0.05 pmol/L (Y/N)?
2. Non-supressed level at week 24, defined as level ≥0.05 pmol/L (Y/N)?
3. Number of weeks for which level ≥0.05 pmol/L - a LOCF approach will be used to calculate number of weeks.

3.2.2.3 Thyroid volume

To determine if daratumumab changes thyroid size
Change in thyroid volume from baseline to 24 weeks.
Thyroid volume assessed at baseline and 24 weeks.

Thyroid volume (ml) is calculated as Height (cm) x Width (cm) x Depth (cm) x 0.479 for the left and right lobe separately. Total volume is the sum of left and right volumes.

The outcome measure will be percentage change from baseline to week 24 in total thyroid volume (ml).

WHO. Assessment of iodine deficiency disorders and monitoring their elimination: A guide for program managers [online]. 3rd edition. Geneva: WHO, 2007. Page 65. Also available from < https://www.who.int/publications/m/item/salt-iodization--assessment-of-iodine-deficiency-disorders-and-monitoring-their-elimination>. ISBN 978 92 4 159582 7

3.2.2.4 Serum ATPO and thyroglobulin antibodies

To determine if daratumumab changes other thyroid autoantibodies
Change in serum ATPO and thyroglobulin antibodies from baseline to 6, 12 and 24 weeks

Serum ATPO (kU/L) and TgAb (IU/ml) are assessed at baseline, weeks 6, 12 and 24 weeks.
Normal levels as specified in Appendix, section 9.1, are:

TPO	<35	kU/L
TgAb	<20	IU/ml

Note there are many instances in the database for which TgAb is recorded as "<20 IU/ml" as a text field validation override.

The outcome measure used for TgAb is the percentage change in TgAb level (IU/ml) from baseline to week 12 and week 24.

The outcome measure used for ATPO is the percentage change in ATPO level (kU/L) from baseline to week 12 and week 24.

For some participants with high levels of ATPO the result is recorded as ">600 kU/L" as a text field validation override. These results will be treated as 600 kU/L.

3.2.2.5 Clinical Activity Score (CAS)

To determine if daratumumab improves thyroid eye disease
Change in CAS score from baseline to 6, 12 and 24 weeks

The Clinical Activity Score is composed of an Objective score (0-5) and a Subjective score (0-2) to produce a combined score ranging from 0-7. For each item, the worst eye is assessed. A high total score indicates worse eye disease.

Proportion with CAS>=3 at baseline, 6, 12 and 24 weeks.

3.2.2.6 Composite eye index

To determine if daratumumab improves thyroid eye disease
Change in composite eye index from baseline to 6, 12 and 24 weeks

The Composite eye index depends on:

- Lid aperture (mm) OD (oculus dexter, right eye) and OS (oculus sinister, left eye)
- Exophthalmos (mm) OD and OS
- Clinical Activity Score (Objective) (5 item)
- Diplopia (Bahn-Gorman) – [0 None, 1 Inconstant, gaze evoked, 2 Intermittent in primary gaze, 3 Constant in primary gaze]
- Dysthyroid Optic Neuropathy(DON) [Y/N]

The criteria for significant improvement are:

a. Reduction in Lid aperture ≥2mm
b. Reduction in Exophthalmos ≥2mm
c. Reduction in objective CAS (5 item) of ≥1
d. Any reduction in Bahn-Gorman score

A ***response*** to treatment is considered positive when there is an improvement in at least 2 of the four features above (a to d) in 1 eye, without concomitant deterioration in the other eye. Deterioration would be defined by the occurrence of DON or worsening of at least 2 of the 4 components a to d.

3.2.2.7 Graves' ophthalmopathy quality of life questionnaire (GOQoL) score

To determine if daratumumab improves thyroid eye disease
Change in GOQoL score from baseline to 6, 12 and 24 weeks.

Following the proposed scoring scheme of Terwee et al *(Interpretation and validity of changes in scores on the Graves' ophthalmopathy quality of life questionnaire(GO-QOL) after different treatments - Clinical Endocrinology, 2001: 54, 391-398)* all GO-QOL questions were scored as `severely limited' (1 point), `a little limited' (2 points), or `not limited at all' (3 points). Questions 1-8 and questions 9-16 were added up to two raw scores from 8 to 24 points, and then transformed to two total scores from 0 to 100 by the following formula:

total score = [(raw score - 8)/16] x 100

For both total scores, higher scores indicate better health. For questions 1 and 2 the answers `no drivers' license' or `never learned to ride a bike' were scored as a missing value. When there were missing values for some items, total scores were calculated for the remaining completed items. The transformation was then adjusted to:

total score = [(raw score - N)/(2 x N)] x 100)

where N is the number of completed items. However, when more than half of the items within a subscale were missing a total score was not calculated but resulted in a missing value.

Terwee et al consider a mean change of at least six points in any of the GO-QOL scores an important change for patients.

3.2.2.8 ThyPRO39 score

Change in ThyPRO39 score from baseline to 6, 12 and 24 weeks

The ThyPro39 score is calculated according to the method in the Appendix, section 10.2

The ThyPro39 consists of 13 separate scales, however for the analysis of the Graves-PCD trial, interest centres on the **Goitre**, **Hyperthyroid**, and **Eye Symptoms** scales and the **Composite** scale.

The items used for the subscales of interest are:

Goitre symptoms scale: 1A, 1C, 1H
Hyperthyroid symptoms scale: 1L, 1M, 1N, 1T
Eye symptoms scale: 1W, 1X, 1BB

The Composite scale is based on 22 items from the Tiredness, Cognition, Anxiety, Depressivity, Emotional Susceptibility, Impaired Social life, Impaired Daily Life and Overall QoL-impact scales: 2A, 2C, 3B*, 4A, 4B, 4F, 5B, 5C, 5E, 6A, 6E, 6G*, 7C, 7D, 7H*, 8A, 8B, 8C, 9A, 9C, 9E, 12. *Positively worded items are scored reversely when constructing scales.

3.2.2.9 Immunoglobulins, specific antibodies including (SARS-CoV2) and blood count

Change in serum immunoglobulins, specific antibodies including (SARS-CoV2) and blood count parameters from baseline to 6, 12 and 24 weeks.

IgG, IgA and IgM levels at baseline, 6, 12 and 24 weeks
Tetanus and Pneumococcus levels and SARS-CoV2 (pos/neg) at baseline, 12 and 24 weeks
Haemoglobin, Platelet, Neutrophil, Lymphocyte and White Cell count at each visit.

Outcome measures will be:

1. Percentage change in level (IgG, IgA and IgM) from baseline to week 6, 12 and 24.
2. Below lower limit of normal at weeks 12 and 24 (Y/N) - as specified in Appendix, section 10.1.

Note that some IgA data is recorded as "<0.04 g/L" in a text field validation override.

Parameter	**Normal range**	**Units**
IgA	0.8 – 3	g/L
IgG	6 – 16	g/L
IgM	0.4 – 2.5	g/L
Tetanus	0 – 40	IU/ml
Pneumococcus	0 – 100	mg/L

3. Pneumococcal antibody > 20 mg/L (Y/N)
4. Tetanus antibody < 0.1 IU/ml (Y/N)

3.2.2.10 Adverse Reactions to 24 weeks

To determine if daratumumab is safe in this patient group

The relationship between the use of IMP/NIMP and the occurrence of each AE is assessed by the PI or delegated clinician using clinical judgement to determine the causal relationship.

Yes (related)	The event is considered related to the IMP/NIMP
Probable	It is probable that the event is related to the IMP/NIMP
Possible	It is possible that the event is related to the IMP/NIMP
Unlikely	It is unlikely that the event is related to the IMP/NIMP
No	The event is not considered related to the IMP/NIMP
Unable to Determine	After review of the information the PI/delegated clinician is unable to determine if the event is related to the IMP/NIMP or not

If the participant is on placebo and the event is considered associated with the placebo (e.g. a reaction to an excipient or impurity within the formulation) such cases are also be assessed as related.

Adverse reactions are considered to be adverse events for which the causality is assessed to be “Yes (related)”, “Probable” or “Possible”.

3.2.3 Dose of IMP received (mg/kg)

Cumulative Daratumumab received.

The prescribed dose of IMP (mg) is administered as an infusion made up of either a 1L or 0.5L saline solution plus a volume of IMP (ml) determined using the Dose Banding Tables in Appendix 3 of Graves-PCD Protocol v6.0, dated 18 May 2023. The volume of IMP (ml) added is dependent on randomised group (mg/kg) and patient weight (kg). IMP is supplied as daratumumab 20mg/ml vials. The saline solution is always 1L for the first treatment. For the second treatment, the saline solution is 0.5L unless the participant experienced a clinically significant infusion related reaction during the first infusion, in which case, a 1L solution is used.

If 100% of the prescribed infusion volume is successfully administered, as is usually the case, then

$$\textbf{\textit{IMP received (mg/kg)}} = \left[\frac{\text{banded dose (mg)}}{\text{patient weight (kg)}}\right].$$

If, however, less than 100% is administered, then

$$\textbf{\textit{IMP received (mg/kg)}} = \left[\frac{\text{percentage of infusion volume administered}}{100}\right] \times \left[\frac{\text{banded dose (mg)}}{\text{patient weight (kg)}}\right]$$

where,

percentage of infusion volume administered =

$$\frac{100 \text{ x } [\text{approximate infusion administered (ml)}]}{[\text{saline solution voulume } (1000 \text{ or } 500\text{ml}) + \text{prescribed IMP volume (ml)}]}$$

and

prescribed IMP volume (ml) = banded dose (mg) / 20 (mg/ml).

3.2.4 Dose of NIMP received

Cumulative prescribed dose of carbimazole or propylthiouracil (PTU) received during the study (baseline to 24 weeks) according to IMP dose. Doses of propylthiouracil will be divided by 10 to analyse as equivalent to carbimazole dose (e.g. 50mg PTU has equivalent potency to 5mg carbimazole). Dosing details of carbimazole and propylthiouracil (PTU) are obtained from thyroid concomitant medication eCRF.

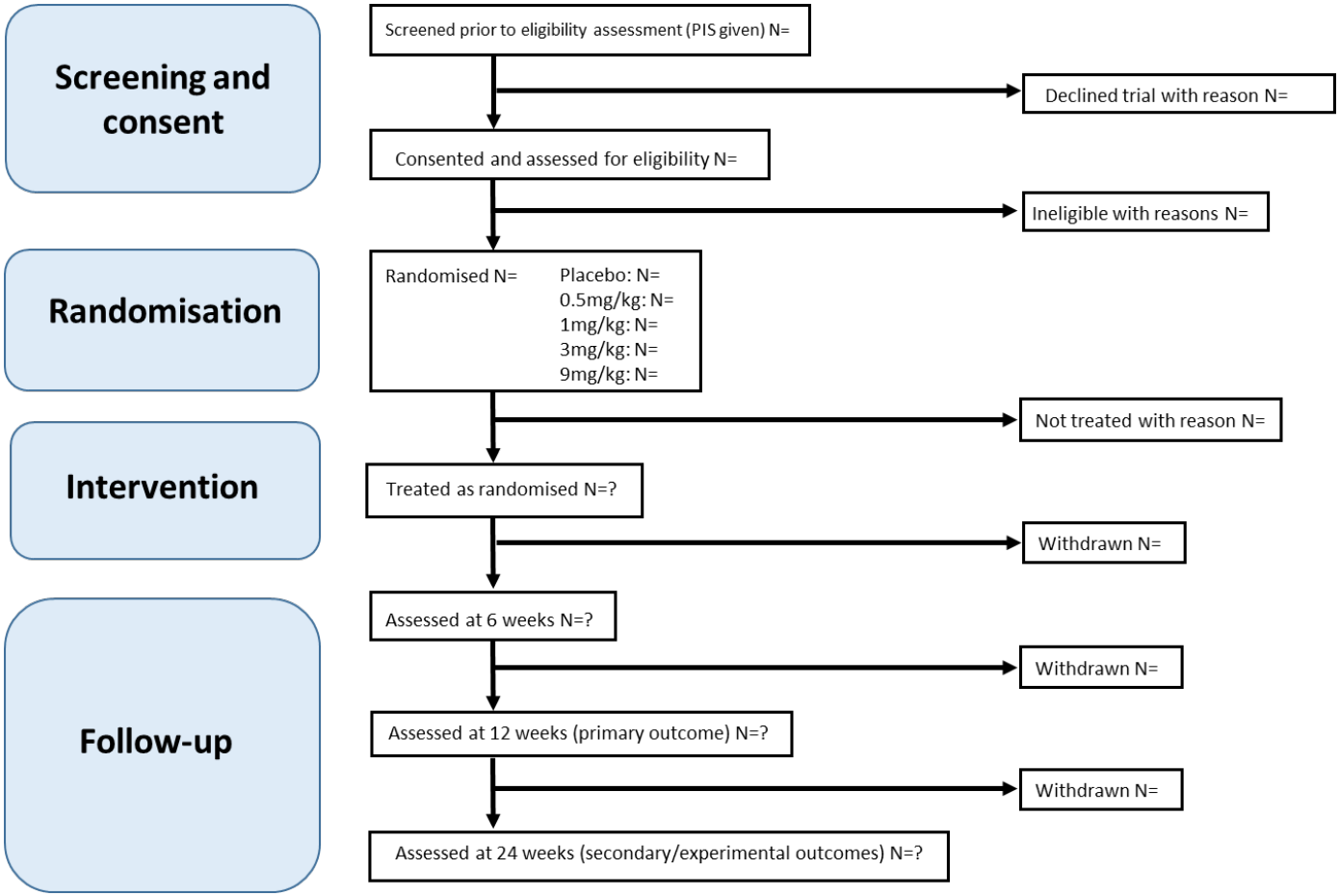
4 Trial Population

4.1 Patient flow through trial

Patient flow through the trial will be presented using a CONSORT diagram. Information will be provided on numbers and reasons (where appropriate) for: patients invited but not interested; consented patients not being eligible; randomised patients subsequently found to be ineligible; patients deviating from treatment as randomised; patients not evaluable for the primary endpoint; patients withdrawing from the trial; patients withdrawing consent and all protocol violations.

The number of ineligible patients and reasons for ineligibility will be reported.

4.2 Example CONSORT flow diagram



4.3 Recruitment

Observed recruitment will be presented graphically over time.

4.4 Analysis populations

Intention-to-treat (ITT) population	Participants analysed according to randomisation allocation following the intention-to-treat principle.
Per-protocol (PP) population	Participants analysed according to whether 75% or more of the full allocated doses of daratumumab or placebo were received.
Dose-response (DR) population	Participants analysed according to the amount of daratumumab received, according to the agreed optimal dose(s) from the interim analysis.
Safety population (SP)	All recruited participants who received any study treatment, regardless of whether it was the protocol specified treatment.

4.5 Baseline characteristics

Intention-to-treat (ITT) population

Demographic, clinical and baseline characteristics at recruitment will be summarised descriptively. For categorical variables, the frequency and percentage in each group will be reported and for continuous variables the mean, standard deviation (SD) and/or median, IQR and range will be reported.

Reported characteristics at Baseline will include:

- Age at randomisation (years),
- Gender (Male/Female)
- Smoking and vaping status
- Current thyroid disease (FT4, FT3, TSH, TRAb at diagnosis)
- Treatment given for Current thyroid disease
- Height (cm), Weight (kg), Pulse (bpm), BMI SD score
- Goitre size (neither palpable nor visible; palpable only, not visible; palpable and visible; large goitre and hence easily seen at a distance)
- Blood Count: Haemoglobin (g/L), Platelets (10^9/l), Neutrophils (10^9/l), Lymphocytes 10^9/l), White Blood Cells (cell/µl or (10^9/l))
- Liver function: ALT (U/L), Bilirubin (µmol/L), Alkaline Phosphatase (IU/L)
- Blood Sampling - IgGAM: IgA, IgG, IgM (g/L)
- Blood sampling - ATPO & TgAb: TPOAb (kU/L), TgAb (IU/ml)
- Thyroid function: TSH (mU/L), FT3 (pmol/L), FT4 (pmol/L), TRAb (U/L)
- Thyroid eye disease (Y/N) - defined as Clinical Activity Score (CAS) ≥3
- Time since diagnosis (defined as date of first thyroid function test showing hyperthyroidism (FT4 and TSH)

Characteristics will be tabulated by randomised dose-group.

4.6 Treatment received

4.6.1 Daratumumab

Intention-to-treat (ITT) population

Treatment received - IMP [Randomised dose (mg/kg), dose of IMP received (mg/kg)] and details of infusion related reactions (IRRs) that led to IMP interruption [Y/N] will be reported by participant, as in table below.

Participant number	Randomised group (mg/kg)	Prescribed dose of IMP (mg/kg)	Dose of IMP received (mg/kg)	Infusion related reactions with IMP interruption (Y/N)
01-003				
01-034				
01-062				
01-001				
01-006				
01-061				

Number and percentage of participants experiencing clinically significant IRRs will be reported by randomised dose group. Clinically significant infusion related reactions are defined as those IRRs which resulted in an interruption to the infusion of IMP.

4.6.2 ATD

Intention-to-treat (ITT) population

Compliance with Anti-Thyroid Medication since last visit [<50%, 50-80%, >80%] recorded at each visit will be tabled by randomised treatment group.

4.6.3 Thyroid concomitant medications

Intention-to-treat (ITT) population

A line listing by participant, ordered by randomised treatment group, will report all thyroid concomitant medication received during the course of the trial.

Randomised group	Participant identifier	Medication	Dose	Dose Units	Frequency	Medication start date	Continuing?	Medication end date
Placebo	01-003	Levothyroxine						
Placebo	01-003	Propanolol						
Placebo	01-003	Propanolol						
Placebo	01-003	Levothyroxine						
Placebo	01-003	Carbimazole						
0.5 mg/kg								
0.5 mg/kg								
0.5 mg/kg								
0.5 mg/kg								
0.5 mg/kg								
1 mg/kg								
1 mg/kg								
1 mg/kg								
1 mg/kg								
1 mg/kg								

4.6.4 Non-thyroid concomitant medications

Intention-to-treat (ITT) population

A line listing by participant, ordered by randomised treatment group, will report all non-thyroid concomitant medication received during the course of the trial.

Randomised group	Participant identifier	Medication	Dose	Units	Frequency	Route	Reason	Start date	Continuing?	Stop date	AE Treatment?
Placebo	01-003										
Placebo	01-003										
Placebo											
Placebo											
Placebo											
Placebo											
Placebo											
0.5 mg/kg											
0.5 mg/kg											
0.5 mg/kg											
0.5 mg/kg											
0.5 mg/kg											
0.5 mg/kg											
1 mg/kg											
1 mg/kg											
1 mg/kg											
1 mg/kg											
1 mg/kg											

4.7 Protocol deviations

Non-administrative Protocol deviations will be provided in a line listing.

5 Interim Analyses, Data Monitoring And Stopping Guidelines

The trial will be monitored by an external DMC that will meet at the start of the trial, for the interim analysis and annually throughout the recruitment and follow-up period of the trial, and on ad hoc basis if required.

The trial includes an interim analysis of the dose-response after 15 participants (n≅3 per allocation) have provided 12-week follow-up data. The interim analysis will consider both efficacy and safety of the daratumumab doses in stage 1. The recommended dose from the stage 1 interim analysis will be presented to the DMC for approval.

During the period of the trial where stage 1 participants are followed-up for 12-week outcome data, it is anticipated that recruited participants may continue to be allocated according to stage 1 allocation until the interim analysis is conducted, and its results are implemented.

No formal statistical stopping rules will be used.

5.1 Dose-response modelling

The analysis population will be the ***Dose-response (DR) population*** (participants analysed according to the amount of daratumumab received).

In order to assess efficacy, a three-parameter Emax model will be fitted to the percentage change in TRAb concentration, defined in 3.2.1. Analysis of the percentage reduction in TRAb will be performed on a log transformed variable δ_t, where

$$\delta_t = -\log_e\left[1 - \frac{PR_t}{100}\right] = \log_e \mathrm{TRAb}^0 - \log_e \mathrm{TRAb}^t \quad (1)$$

i.e. the change from baseline to time *t* in $\log_e$TRAb. Normally distributed data and homoscedastic variance will be assumed.

The Emax model to be fitted is:

$$\delta_t = E_0 + E_{max}\left(\frac{dose}{ED_{50}+dose}\right) \quad (2)$$

where,

dose is the estimated amount of IMP (mg/kg) received by the patient - as defined in section 3.2.3., E_{max} is the asymptotic maximum response attributable to the drug, E_0 is the response when *dose* is 0, and ED_{50} is the *dose* giving response $E_0 + \frac{1}{2}E_{max}$.

The asymptotic plateau on the response scale (i.e. change in $\log_e$ TRAb) is given by

$$E_0 + E_{max} \quad (3)$$

and the dose *D* which gives a response δ_t is given by

$$D = \frac{(\delta_t - E_0)ED_{50}}{E_{max} - (\delta_t - E_0)} \quad (4)$$

Given estimates of E_0, ED_{50} and E_{max} from model (1) the following additional estimates may be obtained:

A. the plateau effect, in terms of PR, by substituting $E_0 + E_{max}$ for δ_t in equation (1)

$$100(1 - exp[-(E_0 + E_{max})])$$

B. the dose that gives 90% of the plateau effect, by substituting $\delta_t = 0.9(E_0 + E_{max})$ into equation (4), and

C. the dose that gives a mean reduction of 50%, by substituting

$$\delta_t = -\log_e\left[1 - \frac{50}{100}\right] = -\log_e(0.5)$$

into equation (4).

It is feasible that TRAb concentrations at week 12, and at other times points, might be below the limit of quantification of the assay. If this situation arises, such values will be replaced by $LOQ/\sqrt{2}$ where LOQ is the limit of quantification of the assay. The use of $\mathrm{LOQ}/\sqrt{2}$ is based on the assumption that data below the LOQ follows a triangular distribution with median $\mathrm{LOQ}/\sqrt{2}$ (i.e. values drop linearly to zero density at zero concentration).

Dose-response modelling will be performed using:

i. change in TRAb from baseline to 12 weeks, and
ii. change in TRAb from baseline to 6 weeks.

5.2 Model-based interim analysis dosing decisions

Model-based stopping criteria and Stage 2 dose recommendations are based on the change in TRAb from baseline to 12 weeks.

1. If the estimated plateau effect (i.e. the maximum percentage reduction in TRAb as the dose of daratumumab increases, A. in 4.3.1) corresponds to >50.0% reduction at 12 weeks, the trial will continue with placebo and the doses B. and C. defined in 4.3.1. Dose B is the dose expected to yield 90% of the estimated maximum TRAb reduction, and C is the dose expected to yield a 50% reduction in TRAb.

If either of the doses B. or C. are close to a stage 1 dose, consideration will be given to proceeding to stage 2 with the approximate stage 1 dose rather than the exact dose defined in 4.3.1.

2. If the estimated plateau effect (A. in 4.3.1) corresponds to <25.0% reduction in TRAb at 12 weeks, the trial will terminate early for lack of promising dose-response unless the DMC agree that there has been an unequivocal improvement in serum thyroid hormone measurements in the active IMP vs placebo groups.

3. If the estimated plateau effect (A. in 4.3.1) corresponds to between a 25%-50% reduction in TRAb at 12 weeks, i.e. 25.0% ≤ plateau effect ≤ 50.0%, the trial will continue with placebo plus doses of 9mg/kg and a higher daratumumab dose (e.g. 12 or 16mg/Kg) as determined by safety data.

5.3 Dosing decisions, other considerations

In addition to the efficacy analysis detailed 4.3.1, the DMC will be provided with line listings of all grade 3 or higher AEs with corresponding dose allocations. This safety information will be taken into consideration in deciding stage 2 dose(s). For instance, if more than one IMP-related severe adverse reaction is observed at higher doses (9 or 3mg/kg), then lower doses will be selected for stage 2 of the trial provided there is evidence for efficacy. The DMC will be guided by the principle that in the absence of any observed dose-related safety signals the dose(s) will be selected for stage 2 according to the three rules stated in 4.3.2.

5.4 Contents of interim analysis report

5.4.1 Dose response modelling results

Dose-response (DR) population

1. Summary of individual IMP dosing and TRAb results – see below
2. Line listing of TRAb results by participant – appendix
3. Plot of TRAb levels versus visit, by individual – highlighting randomised group
4. Plot of percentage reduction in TRAb from baseline versus visit, by individual – highlighting randomised group
5. Summary statistics for percentage reduction in TRAb from baseline to week 6 and 12 by randomised group (mean, median, SD, min, max)
6. Summary statistics for change in $\log_e$TRAb from baseline to week 6 and 12 by randomised group (mean, median, SD, min, max)
7. Plot of change in $\log_e$TRAb from baseline to week 12 versus dose of IMP received (mg/kg) with fitted Emax model
8. Plot of percentage reduction in TRAb from baseline to week 12 versus dose of IMP received (mg/kg) with fitted Emax model
9. Parameter estimates from Emax model (week 12): E_0, ED_{50} and E_{max}
10. The plateau effect, in terms of percentage reduction (week 12)
11. The plateau effect, in terms of change in $\log_e$ TRAb (week 12)
12. The dose that gives 90% of the plateau effect (week 12)
13. The dose that gives a mean reduction of 50% (week 12)
14. Items 7 – 13 repeated for week 6 outcome data
15. Recommended doses for stage 2 based on week 12 results

Summary of individual IMP dosing and TRAb results

Patient number	Randomised group	Dose of IMP received (mg/kg)	TRAB (U/L) at Baseline	TRAB (U/L) at week 6	TRAB (U/L) at week 12	PR in TRAb at week 6	PR in TRAb at week 12

5.4.2 Individual participant details

The DMC will be provided with the following information, by participant and at available visits (see tables below), to provide context to the analysis and to inform the dosing decision. Participant details will be listed by their randomised group. Abnormal blood values will be highlighted – see Appendix section 9.1 for normal ranges.

1. Demographics (gender at birth, age at randomisation)
2. Thyroid disease at diagnosis (age, FT4, FT3, TSH, TRAb)
3. Randomised group
4. TSH (mU/L), FT4 (pmol/L), FT3 (pmol/L), TRAb (U/L)
5. Full blood count (FBC) [haemoglobin (g/L), platelets (10^9L), white cell (10^9L) , neutrophils (10^9L), lymphocytes (10^9L)]
6. Liver function tests (LFT) [ALT (U/L), bilirubin (µmol/L), alkaline phosphatase (IU/L)]
7. U&E test results
8. IgGAM [IgA, IgG, IgM (g/L)]
9. Treatment received - IMP [Randomised dose (mg/kg), prescribed dose (mg/kg), dose received (mg/kg)]
10. Thyroid concomitant medications
11. Non-thyroid concomitant medications
12. Adverse events – grade 3 or higher

Availability of blood results

Parameter	Baseline/ Treatment 1	Treatment 2 (Week 2)	Safety Visit (Week 4)	Early outcome visit (Week 6)	Primary outcome visit (Week 12)
TSH, FT4, FT, TRAb	Yes	Yes	Yes	Yes	Yes
FBC	Yes	Yes	Yes	Yes	Yes
LFTs and U&E	Yes	Yes	Yes	Yes	**No**
IgGAM	Yes	**No**	**No**	Yes	Yes

Dummy table for Safety bloods (e.g. FBC, LFTs, IgGAMs and U&Es)
Values above the normal range shown in red, those below the normal range shown in blue

Participant ID	Randomised group	Visit	Haemoglobin (g/L)	Platelets (10^9/L)	Total White Cell (10^9/L)	Lymphocytes (10^9/L)	Neutrophils (10^9/L)
01-006	Placebo	Visit 4 - Baseline & Treatment 1	123	123	4	2	2
01-006	Placebo	Visit 5 - Treatment 2	123	111	1413	1911	2
01-006	Placebo	Visit 6 - Early Safety Visit	123	115	5011	149	2
01-006	Placebo	Visit 7 - Early Outcome & Safety Visit	123	123	415	13	2
01-006	Placebo	Visit 8 - Primary Outcome Visit	123	100	325	3	2
01-003	1mg/Kg	Visit 4 - Baseline & Treatment 1	121	134	1210	249	1
01-003	1mg/Kg	Visit 5 - Treatment 2	200	215	2	1	1
01-003	1mg/Kg	Visit 6 - Early Safety Visit	121	163	18	1	7
01-003	1mg/Kg	Visit 7 - Early Outcome & Safety Visit	134	222	923	122	21
01-003	1mg/Kg	Visit 8 - Primary Outcome Visit	142	76	817	124	13
01-004	9mg/Kg	Visit 4 - Baseline & Treatment 1	123	541	1622	11	11
01-004	9mg/Kg	Visit 5 - Treatment 2	132	231	85	122	3
01-004	9mg/Kg	Visit 6 - Early Safety Visit	142	132	1819	7	12
01-004	9mg/Kg	Visit 7 - Early Outcome & Safety Visit	111	124	7	12	1
01-004	9mg/Kg	Visit 8 - Primary Outcome Visit	121	143	710	183	7

6 STATISTICAL CONSIDERATIONS

6.1 Timing of analyses

The final analysis will take place after the last recruited participant has attended their 24-week follow-up visit, data queries have been resolved, and the database has been locked. Final analysis is expected to take place in Summer 2024.

6.2 Analysis Methods

6.2.1 Analysis of primary outcome

Using the ITT population (participants analysed according to randomisation allocation), the primary analysis will assess whether there is a significant dose-response using percentage reduction in serum TRAb concentration from **baseline to 12 weeks**, as defined in section 3.2.1. Analysis of percentage reduction in TRAb will be performed on the log transformed variable δ, where

$$\delta = -\log_e\left[1 - \frac{PR_{12}}{100}\right] = \log_e \mathrm{TRAb}^0 - \log_e \mathrm{TRAb}^{12} \quad (1)$$

i.e. the change from **baseline to week 12** in $\log_e$TRAb.

We will report:

1. Summary of individual IMP dosing and TRAb results – see table below
2. Line listing of TRAb results by participant – in an Appendix
3. Plot of TRAb levels versus week number, by individual – highlighting randomised group
4. Plot of percentage reduction in TRAb from baseline versus week number, by individual – highlighting randomised group
5. Summary statistics for percentage reduction in TRAb from baseline to week 6, 12 and 24 by randomised group (mean, median, SD, min, max)
6. Summary statistics for change in $\log_e$TRAb from baseline to week 6, 12 and 24 by randomised group (mean, median, SD, min, max)

Patient number	Randomised group	Dose of IMP received (mg/kg)	TRAB (U/L) Baseline	TRAB (U/L) week 6	TRAB (U/L) week 12	TRAB (U/L) week 24	PR in TRAb week 6	PR in TRAb week 12	PR in TRAb week 24

PR – percentage reduction

To assess whether there is a significant dose-response we will calculate a contrast-based test statistic with a one-sided 5% type I error rate (Thomas N. Understanding MCP-MOD dose finding as a method based on linear regression. Stat Med 2017; 36:4401-4413).

Let μ_j be the mean level of δ for dose-group j $(j = 0, 1, .., k)$, where μ_0 is the mean for the placebo group. To construct a contrast for establishing dose-response, let c_j be the corresponding contrast coefficients with the condition that $\sum_{j=0}^{k} c_j = 0$.

The null hypothesis (H_0) of no treatment effect versus the one-sided alternative hypothesis (H_A) of significant treatment effect can be written as:

$$H_0: L(\mu) = \sum_{j=0}^{k} c_j \mu_j \leq 0 \quad \text{versus} \quad H_A: L(\mu) = \sum_{j=0}^{k} c_j \mu_j > 0$$

The analysis of the GRAVES-PCD trial includes 5 active dose-groups, plus placebo, (0, 0.5, 1, 3, 9 and 16 mg/kg) hence *k = 5* and the contrast coefficients are:

$$\text{c} = (c_0, c_1, c_2, c_3, c_4, c_5) = (-5, -3, -1, 1, 3, 5)$$

taken from "A proof-of-concept clinical trial design combined with dose-ranging exploration" Xin Wang and Naitee Ting, *Pharmaceut. Statist*. 2012, 11 Page. 405.

The null and alternative hypotheses are then:

$$H_0: L(\mu) = -5\mu_0 - 3\mu_1 - \mu_2 + \mu_3 + 3\mu_4 + 5\mu_5 \leq 0$$

versus

$$H_A: L(\mu) = -5\mu_0 - 3\mu_1 - \mu_2 + \mu_3 + 3\mu_4 + 5\mu_5 > 0$$

The one-sided t-test statistic is given by:

$$T = \frac{\widehat{L(\mu)}}{SE\left(\widehat{L(\mu)}\right)} = \frac{\sum_{j=0}^{5} c_j \bar{x}_j}{SE\left(\sum_{j=0}^{5} c_j \bar{x}_j\right)} = \frac{\sum_{j=0}^{5} c_j \bar{x}_j}{\sigma \sqrt{\left(\sum_{j=0}^{5} \frac{{c_j}^2}{n_j}\right)}},$$

where $\bar{x}_j$ is the sample mean of δ for dose-group *j* and σ is the estimated within dose-group standard deviation. The test statistic has *n-6* degrees of freedom where *n* is the total sample size.

If this test statistic is significant, we will then test whether there is a difference in the mean outcome level between individual doses and placebo using the following linear model. Let δ_{ij} denote δ for participant i in dose-group *j (j=0, 1,..., k)*, then the analysis model is given by:

$$\delta_{ij} = \theta + \tau_j . I(D_i = j) + e_{ij}$$

where $I(.)$ is an indicator function and D_i is a label denoting the dose-group for participant i. For the placebo dose-group, $j = 0$ and $\tau_0 = 0$. e_{ij} are random errors with distribution

$N(0, \sigma^2)$. The parameters τ_j *(j=1,..., k)* represent the treatment effect of dose-group *j* relative to the placebo group.

The primary analysis will test the hypothesis: $H_0: \tau_j = 0\ (j = 1, \ldots, 5)$ *versus* $H_A: \tau_j \neq 0\ (j = 1, \ldots, 5)$. Participants are analysed according to their randomised dose-group, hence there are 5 active dose-groups, 0.5, 1, 3, 9 and 16 mg/kg (k=5), plus the placebo group. The type I error level will be set to 5%. We will report 95% confidence intervals for the difference in mean level between each dose-group and placebo group, first transforming confidence intervals back to the percentage reduction scale (rather than reporting intervals for differences in δ),

i.e. $\mathrm{PR}_j = 100\big(1 - exp\big[-\tau_j\big]\big) \qquad (j = 1, \ldots, 5)$

If there is a statistically significant difference between the five dose-groups and placebo, we will report the associated p-values comparing each individual dose-group with placebo.

6.2.2 Secondary analysis of primary outcome

We will perform a simple test of all placebo participants versus all daratumumab participants using separate linear models that adjust for the following binary predictor variables:

1. Inflammatory eye disease at the eligibility screening visit - defined as clinical activity score (CAS)≥3, or thyroid dermopathy
2. Gender at birth – Male versus Female
3. Smoking status - “Smoked within 1 year/current smoker” versus “Hasn't Smoked within 1 year, Never Smoked”
4. Baseline TRAb ≤50 U/L vs >50 U/L

The model used, an extension to the primary analysis model, is given by:

$$\delta_{ij} = \theta + \tau_j . I(D_i = j) + \pi . I(E_i = 1) + e_{ij}$$

where $E_i = 1$ or 0 denotes the level of the binary predictor variable. δ_{ij} again denotes the change from **baseline to week 12** in $\log_e$TRAb for participant i in dose-group *j (j=0, 1,..., k).*

All participants randomised to active dose-groups (i.e. participants randomised to 0.5, 1, 3, 9 or a higher dose) are analysed together in a single group, hence *k=1*. These secondary analyses will test the hypothesis: $H_0: \tau_1 = 0$ *versus* $H_A: \tau_1 \neq 0$ adjusting for each predictor variable separately. The type I error level will be set to 5% for each test. We will report 95% confidence intervals for the difference in mean level between all active dose-groups and placebo adjusting for each predictor variable; in addition 95% confidence intervals for the difference in mean level between levels of the predictor variables - first transforming the confidence interval back to the percentage reduction scale,

i.e. $\mathrm{PR} = 100(1 - exp[-\tau_1])$ and $\mathrm{PR} = 100(1 - exp[-\pi])$.

6.2.3 Non-linear dose-response models

The analysis population will be the ***Dose-response (DR) population*** (participants analysed according to the amount of daratumumab received). We will fit a three-parameter Emax and a quadratic model to assess dose-response. If these models are clearly inconsistent with the observed data, we may fit one or more additional models. These analyses will help inform selection of the most suitable dose for subsequent trials.

Let δ_i to be the change from **baseline to week 12** in $\log_e$TRAb for participant i. We will fit dose-response models of the form:

$$\delta_i = f(d_i, \beta) + e_i,$$

where e_i are i.i.d. random errors with distribution $N(0, \sigma^2)$ and d_i is the estimated amount of IMP (mg/kg) received by participant *i* - as defined in section 3.2.3.

We will fit the Emax model used in the interim analysis:

$$f(d_i, E_0, ED_{50}, E_{max}) = E_0 + E_{max}\{d_i/(ED_{50} + d_i)\}.$$

The following will be reported:

1. Parameter estimates and 95% confidence intervals: E_0, ED_{50} and E_{max}
2. The plateau effect, in terms of percentage reduction
3. The plateau effect, in terms of change in $\log_e$TRAb
4. The dose that gives 90% of the plateau effect
5. The dose that gives a mean reduction of 50%
6. The Akaike information criterion (AIC)

The following graphs will be included:

1. Plot of change in $\log_e$TRAb from **baseline to week 12** versus dose of IMP received (mg/kg) with fitted Emax model
2. Plot of percentage reduction in TRAb from **baseline to week 12** versus dose of IMP received (mg/kg) with fitted Emax model

We will then fit the quadratic model given by:

$$f(d_i, E_0, \beta_1, \beta_2) = E_0 + \beta_1 d_i + \beta_2 {d_i}^2$$

The following will be reported:

1. Parameter estimates and 95% confidence intervals: E_0, β_1 and β_2
2. The dose that gives a mean reduction of 50%
3. The Akaike information criterion (AIC)

The following graphs will be included:

1. Plot of change in $\log_e$TRAb from **baseline to week 12** versus dose of IMP received (mg/kg) with fitted quadratic model

2. Plot of percentage reduction in TRAb from **baseline to week 12** versus dose of IMP received (mg/kg) with fitted quadratic model

If the Emax and quadratic models are clearly inconsistent with the observed data, we may fit one or more of the following models:

1. $f(d_i, E_0, \beta_1) = E_0 + \beta_1 d_i; \quad \beta_1 > 0$
2. $f(d_i, E_0, \beta_1, \beta_2) = E_0 + \beta_1 log(d_i + \beta_2); \quad \beta_2 > 0$
3. $f(d_i, E_0, \beta_1) = E_0 \exp(d_i/\beta_1)$

6.2.4 Subgroup Analyses

If there is a statistically significant difference in the primary outcome between levels of a binary predictor variable used in the secondary analysis of the primary outcome, section 6.2.2, we will fit a further model including an interaction term between treatment and the significant predictor variable. This further analysis will only be carried out if there is a minimum of 10 participants at each level of the binary predictor variable.

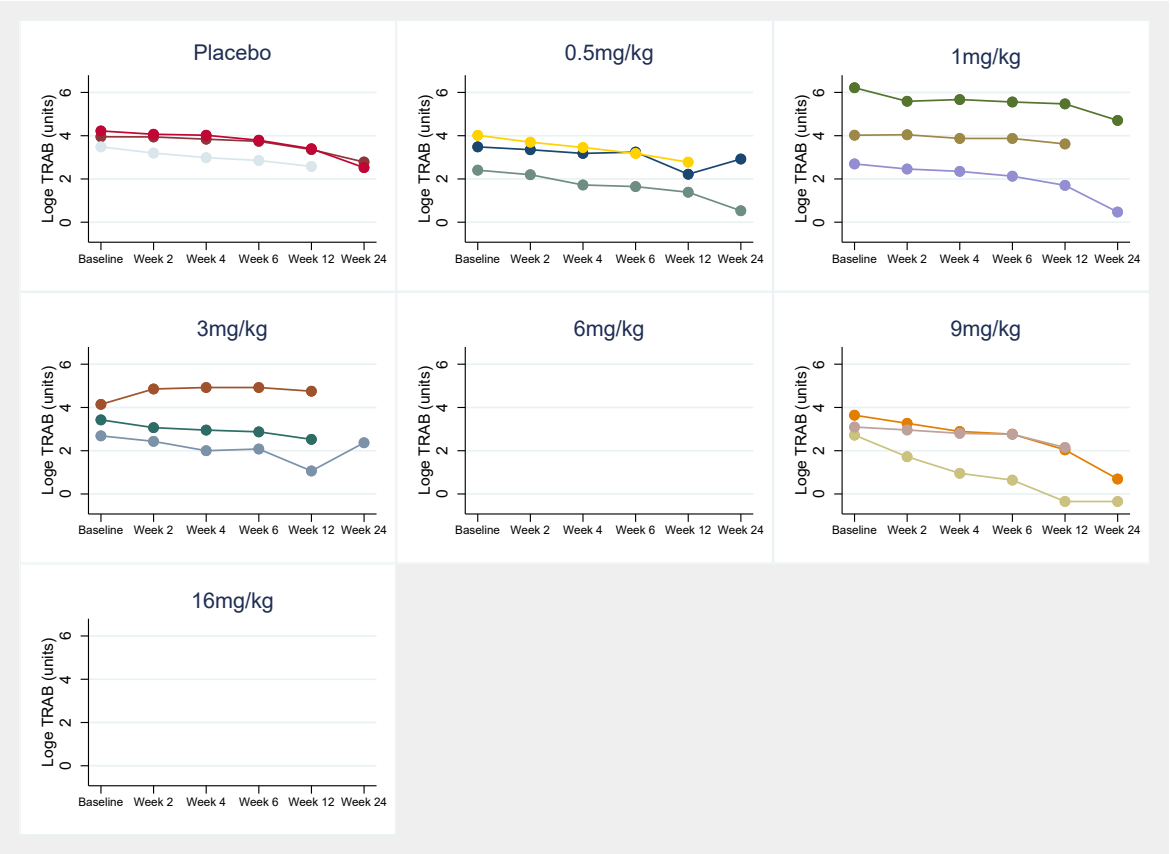
6.2.5 Analysis of secondary outcomes

Using the ITT population (participants analysed according to randomisation allocation), secondary outcome measures will be analysed as either i) continuous, assumed to be normally distributed (possibly after a transformation) or ii) binary, as appropriate. Secondary outcomes are defined in section 3.2.2.

The following sections provide example tables and analyses, the actual randomised dose groups might differ depending on those selected following the Stage 1 interim analysis.

For continuous outcomes (e.g. TRAb, FT3, Thyroid eye volume, etc.), we will provide a plot of outcome level versus week number, by individual – highlighting randomised dose group.

Example graph: Outcome measure by individual and randomised dose group



For derived continuous outcome measures, for example, *percentage change from baseline to week 24*, we will report summary statistics by randomised dose group.

Example table: *Percentage change from baseline to week 24 in total thyroid volume.*

Randomised dose group (mg/kg)	Number of participants	Mean	Median	SD	Minimum	Maximum
0						
0.5						
1						
3						
9						
16						
Total						

For derived binary and categorical outcome measures, for example, *normal FT3 level at week 12 (Yes/No)*, we will report, the number of participants in each category by randomised dose group.

Example table: Number (%) of participants with normal FT3 level at week 12 (Yes/No)?

Randomised dose group (mg/kg)	Yes	No	Total
0			
0.5			
1			
3			
9			
16			
Total			

In addition, for continuous outcome measures (e.g. TRAb, FT3, Thyroid eye volume, etc.), we will report summary statistics by visit and dose group - placebo, low (0.5, 1 and 3mg/kg), high (9m/kg and above).

Example table: Continuous outcomes by visit and dose group (placebo, low, high)

Dose group	Visit	N	Mean	SD	Min	Median	Max
Placebo	Baseline	3					
	Week 2	3					
	Week 4	3					
	Week 6	3					
	Week 12	3					
	Week 24	2					
Low *(0.5, 1 and 3mg/kg)*	Baseline	9					
	Week 2	9					
	Week 4	9					
	Week 6	9					
	Week 12	9					
	Week 24	5					
High *(9mg/kg and above)*	Baseline	3					
	Week 2	3					
	Week 4	3					
	Week 6	3					
	Week 12	3					
	Week 24	2					

6.2.5.1 Statistical analysis

Secondary outcome measures will be analysed as either i) continuous, assumed to be normally distributed (possibly after a transformation) or ii) binary, as appropriate.

The analysis model for continuous outcomes has the same form as the model used for the primary analysis. Let y_{ij} denote a continuous secondary outcome measure for participant i in dose-group *j (j=0, 1,..., k)*, then the analysis model is given by:

$$Y_{ij} = \theta + \tau_j . I(D_i = j) + e_{ij}$$

where, as previously, $I(.)$ is an indicator function, D_i is a label denoting the dose-group for participant i, $\tau_0 = 0$, and e_{ij} are random errors with distribution $N(0, \sigma^2)$.

If Z_{ij} denotes a binary outcome, taking the values 1 or 0 (corresponding to *Yes* or *No*, for example) for participant i in dose-group j $(j = 0, 1, \ldots, k)$, the analysis model for binary outcomes is a standard logistic regression model given by:

$$logit\left[\Pr\left(Z_{ij}\right) = 1\right] = \theta + \tau_j . I(D_i = j)$$

and, as previously, $\tau_0 = 0$.

For both continuous and binary outcomes, participants on active treatment will be grouped into low (0.5, 1 and 3mg/kg) and high (9 mg/kg and above) dose-groups. The parameters τ_j *(j=1, 2)* represent the treatment effect of the low and high dose-groups *relative* to the placebo group.

We will test the hypothesis: H_0: $\tau_1 = \tau_2 = 0$ *versus* H_A: $\tau_j \neq 0$ $(j = 1, 2)$. The type I error level will be set to 5%. For continuous outcomes we will report 95% confidence intervals for τ_j *(j=1, 2)*; for binary outcomes we will report 95% confidence intervals for $exp\left[\tau_j\right]$, i.e. the odds ratios.

Statistical analysis will be carried out using the following derived secondary outcomes measures:

1. Normal FT3 level at week 12, defined as level < UL (Y/N)?
2. Normal FT3 level at week 24, defined as level < UL (Y/N)?
3. Non-supressed TSH at week 12, defined as level ≥0.05 pmol/L (Y/N)?
4. Non-supressed TSH at week 24, defined as level ≥0.05 pmol/L (Y/N)?
5. Percentage change from baseline to week 24 in total thyroid volume (ml).
6. Percentage change in ATPO level (kU/L) from baseline to week 12.
7. Percentage change in ATPO level (kU/L) from baseline to week 24.
8. Change in ThyPRO composite scale score from baseline to week 12
9. Change in ThyPRO composite scale score from baseline to week 24

6.2.6 Additional analyses

6.2.6.1 End of study thyroid treatment outcome

Information relating to thyroid treatment outcome including the decision to treat participants with radioiodine or thyroidectomy surgery, and whether continued use of ATD was required will be reported descriptively only. No hypothesis testing will be performed.

A swimmer plot may be created to visually summarise states such as “On ATD (state dose)”; "Off ATD (Remission)", and endpoints/transitions such as Relapse, RAI therapy, thyroidectomy surgery.

This analysis will be performed outwith the Biostatistics Research Group.

If results from these analyses are reported in the final statistical report, they will be contained within an appendix only. Results from these analyses should only be reported as supplemental material in any trial related publication or report and should not be included in the abstract, or in the discussion. The following statement should be included in any report or publication including these analyses:

Thyroid treatment outcome, specifically, radioiodine use and thyroidectomy surgery, was not a primary or secondary outcome for the GRAVES-PCD trial; data related to these outcomes were collected retrospectively from participant clinical notes at the end of the trial. The data are not stored within the trial CDMS.

7 SAFETY

The following sections provide example tables and analyses by randomised dose group. The actual randomised dose groups might differ depending on those selected following the Stage 1 interim analysis.

An Adverse Event (AE) is defined as any untoward medical occurrence in a participant to whom a medicinal product has been administered, including occurrences which are not necessarily caused by or related to that product.

Adverse reactions (ARs) are defined as an untoward or unintended response in a participant to an investigational medicinal product which is related to any dose administered to that participant. The phrase "response to an investigational medicinal product" means that a causal relationship between a trial medication and an AE is at least a reasonable possibility i.e. the relationship cannot be ruled out. All cases judged by either the reporting medically qualified professional or the Sponsor as having a reasonable suspected causal relationship to the trial medication qualify as adverse reactions.

The relationship between the use of IMP/NIMP and the occurrence of each AE is assessed and categorised by the PI or delegated clinician using clinical judgement to determine the causal relationship. Other factors such as medical history of underlying diseases, concomitant therapy and any other relevant risk factors should be considered. The PI should also consult the current version of the RSI.

Yes (related)	The event is considered related to the IMP/NIMP
Probable	It is probable that the event is related to the IMP/NIMP
Possible	It is possible that the event is related to the IMP/NIMP
Unlikely	It is unlikely that the event is related to the IMP/NIMP
No	The event is not considered related to the IMP/NIMP
Unable to Determine	After review of the information the PI/delegated clinician is unable to determine if the event is related to the IMP/NIMP or not

If the participant is on placebo and the event is considered associated with the placebo (e.g. a reaction to an excipient or impurity within the formulation) such cases must also be assessed as related.

Assessment of the severity of an AE or AR is made according to the table below:

Grade 1	Minor adverse event, not requiring medical intervention. May be asymptomatic and is likely to be a clinical or diagnostic observation only; or may be a symptomatic but minor, or transient event, with no necessity for medical intervention. This might include asymptomatic laboratory or radiographic findings. A minor adverse event is likely to have only marginal clinical relevance.
Grade 2	An adverse event which may require some medical intervention (local/non-invasive) and which is symptomatic to patient. May affect activities of daily living.

Grade 3	Significant symptoms reported, requiring medical intervention, and possibly requiring hospitalisation. Medically significant and likely to be significantly affecting activities of daily living.
Grade 4	An adverse event that requires urgent intervention or may have life-threatening consequences.
Grade 5	Death related to the adverse event.

7.1 Adverse reactions

We will report the number of grade 1 ,2 and 3/4/5 ARs by randomised dose group, and the number of participants with at least 1 ARs by randomised dose group. ***No hypothesis testing will be performed.***

Example table: Summary of adverse reactions (possibly, probably, definitely related) – Safety Population

	Randomised dose group (mg/kg)						**Total**
	0	**0.5**	**1**	**3**	**9**	**16**	
Number of participants	*8*	*3*	*3*	*8*	*8*	*5*	*30*
Number of ARs reported per participant							
0	2						
1	0						
2	3						
3	0						
4+	3						
Mean (SD)	2.5 (1.2)						
Median	2						
Min, Max	0, 6						
Worst grade AR reported per participant							
None	2						
1	5						
2	0						
3, 4 or 5	1						

Example table: Number of ARs by severity and randomised dose group:

Randomised dose group (mg/kg)	Number of participants	Grade			Total
		1	2	3, 4 or 5	
0	8	1	2	2	5
0.5	3				4
1	3				6
3	3				2
9	8				5
16	5				1
Total	**30**				**23**

Example table: Number of participants with at least 1 ARs by randomised dose group:

Randomised dose group (mg/kg)	Number of participants	Number of participants with 1 or more ARs
0	8	
0.5	3	
1	3	
3	3	
9	8	
16	5	
Total	**30**	

A line listing of all adverse reactions ARs will be reported by participant, ordered by randomised dose group

Example table:

Randomised group	Participant number	Adverse Event	Serious (Y/N)	Date of onset	Ongoing	Date of resolution	Relationship to treatment (causality)	Severity	Action taken	AE Outcome
Placebo	01-003									
Placebo	01-003									
Placebo	01-003									
0.5 mg/kg										
0.5 mg/kg										
0.5 mg/kg										
0.5 mg/kg										
9 mg/kg										
9 mg/kg										
9 mg/kg										
9 mg/kg										
9 mg/kg										
9 mg/kg										

7.2 Serious adverse events

For each SAE (regardless of relationship to IMP, i.e. causality) the following information will be reported:

- Full details in medical terms and case description
- Event duration (start and end dates, if applicable)
- Severity
- Action taken
- Outcome
- Seriousness criteria
- Causality in the opinion of the investigator

7.3 Other safety data

We will report the following safety measures/parameters

SARS-CoV2 antibody

The number of participants whose COVID spike antibody became negative following treatment.

Tetanus antibody

The proportion of participants with Tetanus antibody < 0.1 IU/ml at baseline, week 12 and week 24, plus the number of participants with a change to negative (loss of tetanus immunity) at weeks 12 and 24.

Pneumococcus antibody

The number of participants changing from positive to negative pneumococcal antibody status. That is, Pneumococcal antibody > 20 mg/L at baseline with Pneumococcal antibody ≤ 20 mg/L at weeks 12 and at week 24.

Blood parameters

The number of participants developing:

1. Anaemia (Hb < 115g/L in females, <130g/L in males)
2. Thrombocytopenia defined as platelet count <150 10^9L
3. Leukopenia defined as WBC <3.0 10^9L
4. Neutropenia defined as neutrophil count <1.0 10^9L
5. Lymphopenia defined as lymphocyte count <1.0 10^9L

8 STATISTICAL SOFTWARE

Dose-response modelling will be performed using the ***DoseFinding*** package (Bjoern Bornkamp, 2019) in R Statistical Software (v3.6.0; R Core Team, 2019). All other statistical analysis will be performed in Stata V16 (StatCorp, 2020). Statistical analyses will be carried out by the Trial Statistician at the Biostatistics Research Group, PHSI, Newcastle University. All programs will be stored in the School Statistics folder on the PHSI server. A paper master copy of all analysis reports will be stored securely in the statistical section of the trial master file.

9 REFERENCES

(R Core Team, 2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/).

Bjoern Bornkamp (2019). DoseFinding: Planning and Analyzing Dose Finding Experiments. R package version 0.9-17. https://CRAN.R-project.org/package=DoseFinding

10 APPENDIX

10.1 Indicative normal ranges pertinent to Stage 1 interim analysis

Parameter	**Normal range**	**Units**
TSH	0.3 – 4.5	mU/L
FT4	10 – 22	pmol/L
FT3	3.1 – 6.8	pmol/L
TRAb	<1.8	U/L
Haemoglobin	115 – 180	g/L
Platelets	150 – 450	10^9L
Neutrophils	2 – 7	10^9L
Lymphocytes	1.5 – 7	10^9L
White Blood Cells	4 – 11	10^9L
ALT	0 – 40	U/L
Bilirubin	0 – 20	µmol/L
Alkaline phosphatase	20 - 130	IU/L
IgA	0.8 – 3	g/L
IgG	6 – 16	g/L
IgM	0.4 – 2.5	g/L
TPO	<35	kU/L
TgAb	<20	IU/ml
Tetanus	0 – 40	IU/ml
Pneumococcus	0 – 100	mg/L

10.2 Scoring of the ThyPRO-39 questionnaire

Each of the 13 ThyPRO-39 scales is scored as a summary score and transformed to range 0-100. This scoring procedure is described step by step below.

Naming of items
Each item is named TQ{relevant item number}. Itemnumber is indicated left of each item in the questionnaire. Thus, the first item, 'sensation of fullness in the neck' is named TQ1A in the database.

Scoring of responses: Item responses are scored
0 for 'Not at all'
1 for 'A little'
2 for 'Some'
3 for 'Quite a bit'
4 for 'Very much'/'Completely'

Scale content
Each scale consists of the following items:

Goitre symptoms scale: TQ1A TQ1C TQ1H
Hyperthyroid symptoms scale: TQ1l TQ1M TQ1N TQ1T
Hypothyroid symptoms scale: TQ1Q TQ1CC TQ1DD TQ1EE
Eye symptoms scale: TQ1W TQ1X TQ1BB
Tiredness scale: TQ2A TQ2C TQ3B
Cognitive problems scale: TQ4A TQ4B TQ4F;
Anxiety scale: TQ5B TQ5C TQ5E
Depressivity scale: TQ6A TQ6E TQ6G
Emotional Susceptibility scale: TQ7C TQ7D TQ7H
Impaired Social life scale: TQ8A TQ8B TQ8C
Impaired Daily life scale: TQ9A TQ9C TQ9E
Cosmetic Complaints scale: TQ11A TQ11D TQ11E

9a - have difficulty managing your daily life?

9c - not be able to participate in life around you?

9e - feel as if everything takes longer to do?

Category	Have difficulty managingyour daily life	ThyPro9a	N	
Category	Not be able to participate in life around you	ThyPro9b	N	
Category	Feel as if everything takes longer to do	ThyPro9c	N	
Category	Has your thyroid disease affected your	ThyPro11a	N	

Overall QoL: TQ12

Composite scale: TQ2A TQ2C TQ3B TQ4A TQ4B TQ4F TQ5B TQ5C TQ5E TQ6A TQ6E TQ6G TQ7C TQ7D TQ7H TQ8A TQ8B TQ8C TQ9A TQ9C TQ9E TQ12

Scoring the individual scales

The raw scale scores are derived by simply adding the response values (0-4) for all the items in a scale. However, three additional features expand this procedure:

a. reversal of positively worded items,
b. imputation for individual missing item responses and
c. 0-100 transformation.

Reversal of positively worded items

As part of the scoring procedure, items 3B, 6G and 7H have to be reversed, i.e. 'Not at all' scored as 4, 'A little' as 3, 'Some' as 2, 'Quite a bit' as 1 and 'Very much'/'Completely scored as 0.

Imputation for individual missing items

If half or more of the items in a scale is completed, missing items are substituted by the mean of the completed items.

Linear transformation

All scales (except for the Hypothyroid Symptoms, the Overall QoL scale and the Composite scale, see below) are transformed to range 0-100 according to Table 1 below:

Table 1. To the left is the raw score. The corresponding 0-100 score is tabulated for each scale separately. For example, a patient with a raw score on the Goiter Symptoms scale of 6 (e.g. because she answered "Some" to all three Goiter items), will have a 0-100 Goiter Symptoms score of 37. A raw score of 6 on the Tiredness scale, would yield a 0-100 score of 50.

Raw sum score	Final rescaled short-form score										
	Goiter	Hyper	Eye	Tired	Cogni-tion	Anxiety	Depres-sion	Suscep-tibility	Social Life	Daily Life	Appear-ance
0	2	2	1	0	1	1	0	1	0	0	1
1	10	8	8	8	7	10	7	7	8	7	12
2	15	13	14	17	14	18	14	13	17	15	21
3	20	18	20	25	21	26	22	21	25	22	28
4	26	23	25	33	29	34	29	28	33	30	36
5	31	28	32	42	37	41	37	36	42	38	43
6	37	33	38	50	44	49	45	44	50	46	51
7	43	38	45	58	52	56	54	52	58	54	59
8	49	44	52	67	60	63	63	60	67	62	66
9	57	49	60	75	68	71	71	68	75	71	73
10	64	55	68	83	76	79	80	77	83	80	80
11	73	60	78	92	85	87	89	86	92	89	87
12	84	66	89	100	95	96	97	95	100	98	96
13		71									
14		77									
15		84									
16		90									

Transformation of the Hypothyroid Symptoms scale:

The ThyPRO-39 Hypothyroid Symptoms is identical to the original ThyPRO Hypothyroid Symptoms scale and is thus transformed to 0-100 according to the formula

Transformed score=(raw sumscore/16)*100

For example, if a patient answered 'Not at all' to two items, 'A little' to one item and 'Some' to the last item, she would have a raw score of 3 (0+0+1+2). The transformed 0-100 score would then be 3/16*100=19.

Transformation of the Overall QoL-impact scale/item:
The Overall QoL item (TQ12) is rescaled to 0-100 simply by taking the mean raw score and multiply by 25.

Scoring the Composite scale
The Composite scale is based on the 22 items from the Tiredness, Cognition, Anxiety, Depressivity, Emotional Susceptibility, Impaired Social life Impaired Daily Life and Overall QoL scales:

TQ2A TQ2C TQ3B TQ4A TQ4B TQ4F TQ5B TQ5C TQ5E TQ6A TQ6E TQ6G TQ7C TQ7D TQ7H TQ8A TQ8B TQ8C TQ9A TQ9C TQ9E TQ12

The raw score is derived as described above, by summation (with imputation for missing), to range 0-88. The raw score is transformed to 0-100 according to the formula

Transformed score=(raw sumscore/88)*100