PAREXEL International

GlaxoSmithKline

GSK Study 205180 (RENAISSANCE)

A Phase IIa, Double-Blind, Mechanistic Study of GSK3196165 in Combination with Methotrexate Therapy in Subjects with Active Rheumatoid Arthritis Despite Treatment with DMARDs

Reporting and Analysis Plan

PAREXEL Project Number: 227221

SPONSOR SIGNATURE PAGE

Approver	Date	Approval Method
PPD Statistics Leader (Immuno-Inflammation Clinical Statistics)	16-Nov-2017	Email
PPD Programming Manager (Immuno-Inflammation Clinical Programming)	16-Nov-2017	Email
PPD Indication Physician Lead, RA (Immuno-Inflammation Clinical Development)	23-Nov-2017	Email
PPD Clinical Development Director (Immuno-Inflammation Late Stage Clinical Development)	28-Nov-2017	Email
PPD Clinical Development Manager (Immuno-inflammation Global Clinical Sciences and Operations)	20-Nov-2017	Email
PPD Data Quality Lead (Immuno-inflammation Global Clinical Sciences and Operations)	20-Nov-2017	Email
PPD Director (II Clinical Pharmacology Modelling and Simulation)	17-Nov-2017	Email
PPD Safety Development Lead (Global Clinical Safety and Pharmacovigilance)	16-Nov-2017	Email

Clinical Statistics and Clinical Programming Line Approvals:

Approver	Date	Approval Method
PPD Statistics Director (Immuno-Inflammation Clinical Statistics)	21-Nov-2107	Email
PPD Programming Manager (Immuno-Inflammation Clinical Programming)	16-Nov-2017	Email

PAREXEL SIGNATURE PAGE

This document has been approved and signed electronically on the final page by the following:

	Signatory
Author	PPD Project Role: Biostatistics Lead
	Signatory
Reviewer	PPD Project Role: Biostatistics Reviewer
	PPD Project Role: Statistical Programming Lead
	PPD Project Role: MWS Reviewer

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ACPA	Anti-cyclic citrullinated protein antibody
ACR	American College of Rheumatology
ADA	Anti-Drug-Antibody
AE	Adverse event
AESI	Adverse event of special interest
ATC	Anatomical Therapeutic Chemical
BMI	Body Mass Index
CDAI	Clinical Disease Activity Index
CfB	Change from baseline
CI	Confidence interval
CIA	Collagen-induced arthritis
CMC	Carpometacarpal joint
CR	Compliance Ratio
CRP	C-reactive protein
CTCAE	Common Terminology Criteria for Adverse Events v 4.0
DAS28	Disease activity score for 28 different joints
DAS28(CRP)	Disease activity score for 28 different joints with CRP value
DAS28(ESR)	Disease activity score for 28 different joints with ESR value
DCE-MRI	Dynamic contrast enhanced magnetic resonance imaging
DIP	Distal Interphalangeal joint
D_{LCO}	Diffusing capacity or transfer factor of the lung for carbon monoxide
DLCOHCPP	D_{LCO} Hemoglobin-corrected Percent of Predicted
DMARD	Disease modifying antirheumatic drugs
DRE	Disease related Events
ECG	Electrocardiogram
eCRF	Electronic case report form
EOW	Every other week
ESR	Erythrocyte sedimentation rate
EULAR	European League against Rheumatism
EW	Early Withdrawal
F	Figure
FEV1	Forced Expiratory Volume in 1 Second
FVC	Forced Vital Capacity
GM-CSF	Granulocyte-macrophage colony-stimulating factor
GSK	GlaxoSmithKline
HAQ-DI	Health Assessment Questionnaire - Disability Index
IRE	Initial rate of enhancement
ITT	Intent-to-treat
K^{trans}	Exchange rate
LDA	Low disease activity
LLQ	Lower limit of quantification
mAb	Monoclonal antibody
ME	Maximal signal intensity enhancement

MCMC	Markov-Chain-Monte-Carlo
MCP	Metacarpophalangeal joint
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed model repeated measures
MTX	Methotrexate
NRI	Non-Responder imputation
OMERACT	Outcome Measures in Rheumatology
PIP	Proximal interphalangeal joint
PAP	Pulmonary alveolar proteinosis
PD	Protocol deviation
PDS	Protocol deviation specification
PFT	Pulmonary function test
PhGA	Physician's Global Assessment of Arthritis
PK	Pharmacokinetics
PK/PD	Pharmacokinetics/Pharmacodynamics
PRO	Patient reported outcome
PT	Preferred term
PtGA	Patient's Global Assessment of Arthritis Disease Activity
QTcF	Fridericia's Correction Formula
RA	Rheumatoid arthritis
RAMRIQ	Rheumatoid arthritis MRI quantitative score
RAMRIS	Rheumatoid arthritis MRI scoring system
RAP	Reporting and Analysis Plan
RF	Rheumatoid Factor
SAE	Serious adverse event
SAS	Statistical Analysis Software
SD	Standard deviation
SDAI	Simplified Disease Activity Index
SJC	Swollen joint count
SJC28	Swollen joint count for 28 different joints
SJC66	Swollen joint count for 66 different joints
SOC	System organ class
SOP	Standard Operating Procedure
SRT	Safety Review Team
T	Table
STM	Set to Missing
TJC	Tender joint count
TJC28	Tender joint count for 28 different joints
TJC68	Tender joint count for 68 different joints
ULQ	Upper limit of quantification
VAS	Visual analogue scale
V_e	Interstitial volume
VEP	Volume of Enhancing Pannus
V_p	Plasma volume

1 INTRODUCTION

GSK3196165 is a novel human monoclonal anti granulocyte-macrophage colony-stimulating factor (GM-CSF) antibody that is being developed for the treatment of rheumatoid arthritis (RA).

Rationale:
This study is designed to explore the activity of the GM-CSF signaling pathway in subjects with RA, and the potential impact of inhibition of this axis by GSK3196165. An additional exploratory aim is to evaluate whether there are any differences in the GM-CSF axis between subjects with early RA (≤2 years since diagnosis) compared with those with more established disease.
Finally, this study aims to establish the potential impact of GSK3196165 on inflammatory structural joint damage in the hand/wrist using magnetic resonance imaging (MRI).

Rheumatoid Arthritis:
RA is a chronic, systemic inflammatory autoimmune disease, characterized by asymmetrical polyarthritis that is associated with substantial disability and morbidity. RA affects approximately 0.5-1.0% of the worldwide population, primarily women, with a peak incidence of onset between 40 and 60 years of age.
Disease modifying antirheumatic drugs (DMARDs) are the cornerstone of RA treatment throughout all stages of disease, and have been demonstrated to maintain or improve physical function and retard radiographic damage. This wide class of drugs includes conventional DMARDs, of which Methotrexate (MTX) is the gold standard, and biological DMARDs which target cytokines (*e.g.* tumor necrosis factor alpha [TNFα], interleukin [IL]-6), B-cells or T-cells. However, a substantial proportion of subjects either fail to respond, or have inadequate response, to currently available RA therapies. Therefore, there is still a medical need for more effective treatments for RA with alternative mechanisms of action.
Intensive treatment (*i.e.* with a biologic drug) early in the disease course of RA, provides an opportunity to induce a sustained remission that can be maintained on conventional DMARDs alone thereby limiting the overall exposure to biological treatments during a subject's lifetime, which should translate into a better overall safety profile.

GM-CSF and RA:
GM-CSF, in combination with other inflammatory stimuli, can activate macrophages, which produce a range of inflammatory cytokines, such as TNFα, IL-6, IL-1, IL-12p70 and IL-23 and various chemokines, and can also express cell surface major histocompatibility complex (MHC) class II heterodimers and present antigen to T cells, further contributing to the inflammatory process.
Accumulating evidence suggests that the GM-CSF pathway may play a central role in the pathogenesis of RA, via the activation and differentiation of neutrophils and macrophages. GM-CSF induces the proliferation and activation of macrophage lineage cells leading to strongly increased production of key proinflammatory cytokines

(including TNFα, IL-6, and IL-1), chemokines and matrix degrading proteases. GM-CSF also serves as a differentiation factor for dendritic cells and induces upregulation of cell surface MHC class II heterodimers on antigen presenting cells, which in turn will activate $CD4^+$ T cells. In addition, GM-CSF is a strong chemoattractant factor for neutrophils and induces the release of reactivated oxygen species from neutrophils, which can directly damage cartilage structure.

GM-CSF and its receptors are found abundantly in the synovial fluid, synovial tissue and plasma of patients with RA. GM-CSF also contributes to osteoclastic bone resorption and subsequent joint damage. Synovial $CD68^+$ macrophages from RA patients correlate with disease activity scores and are potential biomarkers for treatment response. The number of macrophages in synovial tissue is correlated with radiographic progression. Moreover, there is evidence that early RA is the result of an already-established synovitis in which macrophage-derived cytokines play an important role in the clinical signs of inflammation.

In mouse models of collagen-induced arthritis (CIA), anti-GM-CSF treatment reduced disease activity and prevented progression of established arthritis and, furthermore, administration of recombinant GM-CSF led to exacerbation of arthritis. Importantly, anti-GM-CSF was shown to be beneficial for either early or established arthritis. Moreover, GM-CSF depletion appeared to have a more dramatic effect than depletion of TNF since only 2/15 (13%) of GM-CSF-deficient mice developed CIA compared with 8/26 (31%) of TNF-deficient mice.

Taken together, pre-clinical and clinical data suggest that GM-CSF is a key mediator of inflammatory and immune disorders and central to RA pathogenesis (particularly in the early stages of RA), providing a strong rationale for considering it as a candidate for therapeutic intervention. Blocking GM-CSF should interfere with several pathophysiological pathways and significantly reduce inflammation by inhibiting activation of inflammatory cells and by blocking the chemotaxis of such cells into the joint thus inhibiting bone and cartilage destruction.

GSK3196165:

GSK3196165 is a high-affinity recombinant human monoclonal antibody (mAb) that binds specifically to human GM-CSF and neutralizes its biological function by blocking the interaction of GM-CSF with its cell surface receptor.
Detailed information relating to non-clinical pharmacology, safety pharmacology, pharmacokinetics (PK) and metabolism, toxicology and other pre-clinical and clinical data with GSK3196165 can be found in the GSK3196165 Investigator's Brochure.

Contents of this document:
This Reporting and Analysis Plan (RAP) is based upon the following study documents:

- Study Protocol, Protocol Amendment 2015N261421_01 (01-March-2016)
- electronic Case Report Form (eCRF) Blank Casebook, Version 3.0 (18-Sep-2017)
- Independent Review Charter, Version 2.0 (24-Jan-2017)

2 STUDY OBJECTIVES

Primary objectives	Primary endpoints
The main objectives of this study are to explore the activity of GM-CSF signaling pathway characterized by exploratory biomarkers in subjects with RA, the impact of GSK3196165 therapy, and whether there are any differences in this GM-CSF signaling pathway between subjects with early RA or established RA.	• Changes from baseline in exploratory biomarkers.
Secondary objectives	**Secondary endpoints**
To evaluate the safety and tolerability of GSK3196165 in subjects with RA.	• Incidence of adverse events (AEs), serious adverse events (SAEs), adverse events of special interest (AESIs). • Immunogenicity (anti-drug antibodies [ADAs]).
To evaluate the impact of GSK3196165 on inflammatory structural joint damage in the hand/wrist using MRI.	• Change from baseline in synovitis, osteitis and erosion as assessed by Outcome Measures in Rheumatology (OMERACT) rheumatoid arthritis MRI scoring system (RAMRIS) and rheumatoid arthritis MRI quantitative score (RAMRIQ) in the most affected hand/wrist.
Exploratory Biomarker Endpoints	
• Identify/validate markers of downstream signaling of GM-CSF and/or potential novel markers for GSK3196165 activity. • Pharmacodynamic biomarkers to assess response to GSK3196165 (*e.g.* may include, but not limited to, IL-1β, TNFα, IL-6, IL-8, IL-15, IL-17A/F, sCD163; chemokines such as CCL17, CXCL13, CCL22; and complement proteins such as C5a, TCC, C4, C4a). • Effect of GSK3196165 on target cell populations (*e.g.* may include, but not limited to, circulating levels of T, B, natural killer (NK), Th17, T regulatory cells and activated monocytes, dendritic cells, neutrophils). • Biomarkers of extracellular matrix (ECM) or aggrecan degradation (*e.g.* may include, but not limited to, serum chondrex [YKL-40] also called human cartilage glycoprotein 39 [HC gp 39], ARGS neoepitope). • Biomarkers which may be indicative of RA disease activity (*e.g.*, may include, but not limited to 14-3-3η, ACPA, CRP, RF). • Pharmacodynamic biomarkers which may be predictive of response to GSK3196165 (*e.g.* may include, but not limited to, C1M, C2M, C3M, CRPM, VICM, MRP8/14, matrix metalloproteinase 3 [MMP-3], 14-3-3η).	

Exploratory MRI/Imaging Endpoints	

• Change from baseline in joint inflammation as measured by dynamic contrast enhanced (DCE)-MRI in the most affected hand/wrist:
  • Exchange rate (K^{trans})
  • Interstitial volume (V_e)
  • Plasma volume (V_p)
  • Initial rate of enhancement (IRE)
  • Maximal signal intensity enhancement (ME).	

Exploratory Clinical Efficacy Endpoints	

• Change from baseline at all assessment timepoints for:
  • ACR20/50/70 response rates
  • Disease activity score for 28 different joints with CRP value (DAS28[CRP]) score
  • DAS28(CRP) remission rates and categorical (European League Against Rheumatism [EULAR] good/moderate) response.
Note: For composite endpoints, *e.g.*, DAS28(CRP), ACR response, etc., each component of the assessment will also be reported. Results over time, reflecting all assessment time points, will also be reported (*e.g.*, graphically, as well as in Tables and Listings).	

Pharmacokinetic/Target Engagement Endpoints	
• GSK3196165 PK parameters derived from the sparse blood samples and using a population PK analysis. • Pharmacodynamic biomarkers to assess target engagement (TE) (*e.g.*, serum concentration of free GM-CSF, GM-CSF: GSK3196165 complex).	
Exploratory Safety Endpoints	
• To evaluate potential biomarkers of pulmonary alveolar proteinosis (PAP) pathogenesis.	• Biomarkers which may be predictive of lung damage (*e.g.*, Krebs von den Lungen-6 [KL-6], surfactant protein-D [SP-D], cholestenoic acid). • GM-CSF auto-antibodies.

Definitions of efficacy endpoints can be found in Section 3.3.

3 INVESTIGATIONAL PLAN

3.1 Overall Study Design and Plan

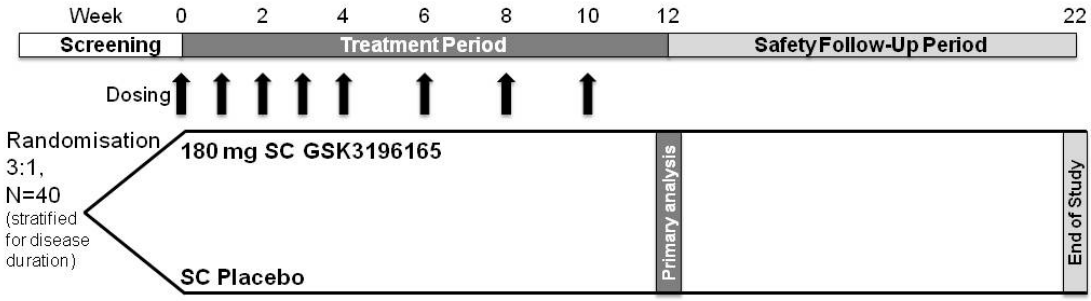
3.1.1 Overall Design

This is a randomized Phase IIa, multi-center, double-blind, placebo-controlled parallel group study to explore the mechanistic evidence that the GM-CSF signaling pathway is active in subjects with RA, and the potential impact of inhibition of this axis by GSK3196165. In addition, the study will assess the potential impact of GSK3196165 on inflammatory structural joint damage in the most affected hand/wrist using MRI.

Following a screening period of up to 6 weeks, approximately 40 subjects with active RA despite treatment with DMARDs (including conventional or biologic) will be randomized in a 3:1 (GSK3196165:placebo) ratio into the study. The total treatment period is up to 10 weeks, with a 12-week follow-up period after the last dose. Ideally, 50% of subjects enrolled into the study should have early RA (*e.g.* ≤2 years since diagnosis). The actual proportion will be monitored throughout the study and randomization may be altered if the projected final proportion with early RA is less than 30%.

A schematic representation of the study design is shown in Figure 1.

Figure 1 Study design



Study Visits

Screening assessments will be performed up to 42 days prior to the baseline visit (Visit 2) on Day 1 (start of study treatment). Subsequent visits are planned on Days 8, 15, 22, 29, 43, 57, 71, 85 (i.e. Week 12), with a follow-up visit in Week 22. Subjects who discontinue the study treatment should have an early withdrawal visit with a subsequent follow-up visit ≥12 weeks after the last dose of study treatment. The exact timing of each assessment is listed in the Time and Events Table 3.1.2.

3.1.2 Time and Events Table

Procedures	Screening - up to 6 weeks	Base-line	Treatment Period								FU	EW[1]
			Week									
			1	2	3	4	6	8	10	12	22	
	Visit											
	1	2	3	4	5	6	7	8	9	10	11	
	Day											
		1	8	15	22	29	43	57	71	85	155	
Written Informed Consent(s)	X											
Subject Demography	X											
Medical, Disease, Therapy History	X											
Inclusion/Exclusion Criteria	X											
PRO[2] and efficacy assessments[3]												
Patient's Assessment of Arthritis Pain, Patient's Global Assessment of Arthritis, Physician's Global Assessment of Arthritis[2]	X	X[4]	X	X	X	X	X	X	X	X	X	X
HAQ-DI[2]	X	X[4]	X	X	X	X	X	X	X	X	X	X
RA Symptom and Impact Diary[2]	X	X[4]	X				X			X	X	X
Swollen (66) & Tender (68) Joint Count[3]	X	X[4]	X	X	X	X	X	X	X	X	X	X
MRI	X[20]					X[21]				X[21]	X[21]	X[22]
Safety Evaluations[5]												
Concomitant Medication	X	X[4]	X	X	X	X	X	X	X	X	X	X
Physical Examination[6]	X	X[4]								X		X
Vital Signs	X	X	X	X	X	X	X	X	X	X		
12-lead ECG[7]	X									X		X
AEs/SAEs[8]	X	X	X	X	X	X	X	X	X	X	X	X
Cough, Lung Auscultation, Pulse Oximetry, Borg Dyspnea Scale	X	X	X	X	X	X	X	X	X	X	X	X
Chest X-ray[9]	X											
Spirometry (FEV1, FVC), D_{LCO}	X[10]									X[11]	X[11]	X[11]

Procedures	Screening - up to 6 weeks	Base-line	Treatment Period								FU	EW[1]
			Week									
			1	2	3	4	6	8	10	12	22	
			Visit									
	1	2	3	4	5	6	7	8	9	10	11	
			Day									
		1	8	15	22	29	43	57	71	85	155	
Laboratory Assessments												
Hematology, Chemistry	X	X[4]		X		X		X		X	X	X
TB, HBsAg, HepB cAb, HepC Ab, HIV	X											
CRP, ESR[12]	X	X[4]	X	X	X	X	X	X	X	X	X	X
RF, ACPA (anti-CCP)		X[4]										
Cholesterol, triglycerides, HDL, LDL[13]		X[4]								X	X	X
Pregnancy test[14]	S	U				U		U		U	U	U
Urinalysis (dip stick)	X					X		X		X	X	X
Other Laboratory Assessments												
PK Sampling (GSK3196165)[15]		X	X	X		X	X	X		X	X	X
GM-CSF & PD blood biomarkers	X	X[4]	X	X		X	X	X		X	X	X
Whole blood flow cytometry		X[4]	X			X				X	X	X
RNA		X[4]								X	X	X
PGx sampling DNA[16]		X[4]										
Lung biomarkers[17]		X[4]								X	X	X
Anti-GM-CSF auto-antibodies[17]		X[4]										
Immunogenicity[18]		X[4]		X		X				X	X	X
Study Treatment GSK3196165/placebo[19] (Methotrexate and folic acid weekly throughout treatment with GSK3196165)		X	X	X	X	X	X	X	X			

Abbreviations: ACPA=anti-cyclic citrullinated protein antibody; AE=adverse event; CCP=cyclic citrullinated peptide; CRP=C-reactive protein; DNA=deoxyribonucleic acid; D_{LCO}=diffusing capacity of the lung for carbon monoxide; ECG=electrocardiogram; ESR=erythrocyte sedimentation rate; EW=early withdrawal; FEV1=forced expiratory volume in one second; FU=follow-up; FVC=forced vital capacity; GM-CSF=granulocyte-macrophage colony-stimulating factor; HAQ-DI=Health Assessment Questionnaire Disability Index; HDL=high density lipoprotein; HBsAg=hepatitis B surface antigen; HepB cAb=anti-hepatitis B core antibody; HepC Ab=hepatitis C virus antibody; HIV=human immunodeficiency

virus; LDL=low density lipoprotein; MRI=magnetic resonance imaging; PD=pharmacodynamic; PGx=pharmacogenomic; PK=Pharmacokinetic; PRO=patient reported outcome; RA=rheumatoid arthritis; RF=rheumatoid factor; RNA=ribonucleic acid; S=serum; SAE=serious adverse event; TB=*Mycobacterium tuberculosis*; U=urine.

1. All subjects who discontinue study medication prematurely should have an early withdrawal (EW) visit as soon as possible after study agent discontinuation and then return for a follow-up safety visit at least 12 weeks after last dose of study medication.
2. All patient reported outcome (PRO) assessments should be conducted before any tests, procedures, assessments or consultations, to avoid influencing the subjects' perception.
3. The same individual (where possible) should perform all disease assessments for an individual subject (with separate joint assessor).
4. Assessments may be performed up to 24h before dosing.
5. All safety evaluations should be conducted before dosing.
6. Brief physical examination (assessments of the skin, lungs, CV system, and abdomen [liver and spleen]) at visits after screening.
7. Electrocardiogram (ECG) should be performed before vital signs, blood draws and dosing (triplicate ECGs required at screening, and single thereafter unless there are safety concerns, in which case repeats may be required (see Section 7.6.10)).
8. AEs/SAEs reported from start of study medication until the last follow-up visit. Any SAEs assessed as related to study participation (*e.g.*, protocol-mandated procedures, invasive tests, or change in existing therapy) or related to a GSK product will be recorded from the time a subject consents to participate in the study up to and including any follow-up contact.
9. Unless performed within previous 12 weeks (no repeat if subject re-screened).
10. If D_{LCO} ≥60% but <70% predicted, a chest HRCT must be performed. If this cannot be done within the screening window, then the subject must be re-screened.
11. Consider repeat D_{LCO} if >15% relative decrease from baseline (recommendations provided in the Pulmonary Safety Guidance document).
12. ESR measured locally.
13. >8h fasting required before blood draw.
14. For women of child-bearing potential. S=serum; U=urine.
15. Blood samples taken before dosing.
16. In consenting subjects.
17. To be analyzed at end of study or in event of pulmonary safety signal.
18. In addition to these scheduled immunogenicity assessments, "event-driven" testing (see Section 7.8) will also be employed for those subjects that experience anaphylaxis, serious hypersensitivity, or adverse events related to study drug administration that led to withdrawal from the study.
19. GSK3196165 or placebo must be administered on the same day each week ±1 day for the first 5 weekly doses, thereafter on the same day every other week (EOW) ±3 days.
20. Subjects must have passed all screening assessments, including laboratory tests, prior to undertaking MRI scanning.
21. MRI may be performed up to 7 days after scheduled visit.
22. MRI should be performed at Early Withdrawal visit if previous MRI was done >14 days before.

3.1.3 Treatment Arms and Duration

Subjects will be randomized on a 3:1 basis to GSK3196165 or placebo. Subjects will receive GSK3196165 180 mg or placebo once weekly from Baseline up to Week 4, and then every other week (EOW) until Week 10.
In addition to GSK3196165 or placebo, subjects will receive MTX 7.5–25 mg/week and folic (or folinic) acid ≥5 mg/week during the combination treatment period.

3.1.4 Study Oversight

Medical monitoring will happen in the form of monthly blinded data review meetings which include but are not limited to:

- Medical review of protocol deviations (PDs)
- Medical review of high panic lab alerts
- Cumulative data review (includes summaries of reason for screening failures, disposition, adverse events (AEs), serious adverse events (SAEs), disease related events (DREs) and chest x-rays)
- Medical review of coding

There will be no pause in recruitment whilst the review of safety data. Further information on the medical monitoring can be found in the Medical Data Review Plan and the Medical Monitoring Plan.

No interim analysis is planned for this study.

The final analysis will be conducted when all subjects have completed their last follow up visit.

3.1.5 Type and Number of Subjects

Approximately 80 subjects with active early/established RA despite treatment with DMARDs will be screened (subjects can be rescreened once) to achieve approximately 40 randomized subjects. Rescreening criteria are listed in the Study Reference Manual (SRM).

3.2 Exploratory Biomarkers

A full list of all biomarkers that will be assessed is given in a summary table in Appendix A (Section 6.1).

3.3 Efficacy and Safety Variables

3.3.1 Disease Activity Score (DAS):

The DAS assessment is a derived measurement with differential weighting given to each component.

The DAS28(CRP) and DAS28(ESR) will be calculated at each assessment time point. The components of the DAS28 arthritis assessment include:

- Tender Joint Count 28 (TJC28)
- Swollen Joint Count 28 (SJC28)
- C-reactive protein (CRP) (in mg/L) or Erythrocyte sedimentation rate (ESR) (in mm/hr)
- Patient's Global Assessment of Arthritis Disease Activity (PtGA) (visual analogue scale with values from 0=best to 100=worst)

DAS28(CRP)

The DAS28(CRP) score will be calculated using the following formula:

$$\text{DAS28(CRP)} = 0.56 * \sqrt{TJC28} + 0.28 * \sqrt{SJC28} + 0.36 * \ln(\text{CRP} + 1) + 0.014 * \text{PtGA} + 0.96.$$

If one of the components is missing at an individual assessment point, the DAS28(CRP) value for that assessment will be set to missing.

DAS28(ESR)
The DAS28(ESR) score will be calculated using the following formula:

$$\text{DAS28(ESR)} = 0.56 * \sqrt{TJC28} + 0.28 * \sqrt{SJC28} + 0.7 * \ln(\text{ESR}) + 0.014 * \text{PtGA}$$

If one of the components is missing at an individual assessment point, the DAS28(ESR) value for that assessment will be set to missing. An ESR value of 0 will be substituted with ESR=1 for the calculation of DAS28(ESR).
DAS28(ESR) will not be summarized or analyzed but will only be included in analysis datasets and listing. Post-hoc analyses may be performed.

DAS28 Remission
DAS28(CRP) remission is achieved by a DAS28(CRP) value lower than 2.6. Similarly, DAS28(ESR) remission is achieved by a DAS28(ESR) value lower than 2.6. Missing DAS28 scores will be considered as not achieving remission.

Categorical DAS28 Response
DAS28(CRP) and DAS28(ESR) scores will each be categorized using EULAR response criteria. Response at a given time point is defined based on the combination of current DAS28 score and the improvement in the current DAS28 score relative to baseline. The definition of no response, moderate response and good response is captured in the following table:

Table 1 EULAR Response Criteria

Current DAS28	DAS28 decrease from baseline value		
	>1.2	>0.6 to ≤1.2	≤ 0.6

≤ 3.2	Good response	Moderate response	No response
> 3.2 to ≤ 5.1	Moderate response	Moderate response	No response
> 5.1	Moderate response	No response	No response

If the post-baseline DAS28(CRP) or DAS28(ESR) score is missing, then the corresponding EULAR category will be missing. If the baseline DAS28(CRP) or DAS28(ESR) is missing, then all post-baseline EULAR response values will be missing.

Categorical DAS28(ESR) will not be summarized or analyzed but will only be included in analysis datasets and listings. Post-hoc analyses may be performed.

DAS28 Low Disease Activity (LDA)
DAS28(CRP) and DAS28(ESR) Low Disease Activity is defined as a DAS28 score of <3.2.
DAS28(ESR) LDA will not be summarized or analyzed but will only be included in analysis datasets and listings. Post-hoc analyses may be performed.

3.3.2 ACR Response Rates

The American College of Rheumatology's (ACR) definition for calculating improvement in RA is calculated as a 20% improvement (ACR20) in both tender and swollen joint counts and 20% improvement in at least 3 of the 5 remaining ACR-core set measures: patient and physician global assessments, patient's assessment of arthritis pain, disability (HAQ-DI), and an acute-phase reactant (i.e. CRP value). Similarly, ACR50 and ACR70 are calculated with the respective percent improvement. This efficacy measurement will be made at every post-baseline study assessment time point.

The specific components of the ACR Assessments that will be used in this study are:

- Tender/Painful Joint count 68 (TJC68)
- Swollen Joint Count 66 (SJC66)
- Patient's Assessment of Arthritis Pain
- Patient's Global Assessment of Arthritis Disease Activity
- Physician's Global Assessment of Arthritis
- CRP (mg/L)
- Health Assessment Questionnaire – Disability Index (HAQ-DI)

For all visits, if any of the component scores are missing, then those scores will be considered as not having met the criteria for improvement. Therefore, if TJC68 or SJC66 or 3 or more of the 5 remaining ACR-core set measures are missing, ACR20/ ACR50/ ACR70 will each be considered as "no response" in the non-responder imputation (NRI) dataset.

For component scores with missing Baseline values or a Baseline value of 0, the percentage improvement can't be calculated and the component will be considered as not having met the criteria for improvement for all visits. If the baseline value is missing, no imputation based on data from the screening visit will be done.

3.3.3 Swollen and Tender/Painful Joint Count

Four different scores will be calculated to evaluate swelling and tenderness of joints. TJC28 and SJC28 will take 28 joints into account, SJC66 and TJC68 will use 66 and 68 joints, respectively.

The assessment for swelling is the total number of joints with a present swelling and ranges from 0 to 28 for SJC28 and 0 to 66 for SJC66.

The assessment for tenderness is the total number of joints with a present tenderness and ranges from 0 to 28 for TJC28 and 0 to 68 for TJC68.

The following 28 joints will be taken into account for TJC28 and SJC28: Shoulder (2 joints), Knee (2), Elbow (2), Wrist (2), Fingers (PIP, MCP: 20).

Additionally the following joints will be taken into account for SJC66/TJC68:
Temporomandibular (2), Sternoclavicular (2), Acromioclavicular (2), Fingers (DIP: 8), Ankle (2), Tarsus (2), Toes (PIP, MTP: 20), Hip (2, only for TJC).
Artificial, ankylosed and missing joints are excluded from swelling and tenderness assessment.

If there are missing or excluded observations for tender or swollen joints then the remaining observations will be assessed and weighted by dividing the number presented by number of non-missing and by multiplying by 28/66/68 for the joint count. No imputations for individual joints will be done. If a joint is not evaluable at Baseline, then that joint is set to missing throughout the study.

If data for more than 50% of the joints are missing at the time of a given assessment, then the total count will be set to missing for that visit.

Observed joint states will be listed without any modification for every subject and visit.

3.3.4 Patient's Assessment of Arthritis Pain

Subjects will assess the severity of their current arthritis pain using a continuous visual analog scale (VAS) with anchors "0" (no pain) and "100" (most severe pain).

No Imputations for missing data will be done.

3.3.5 Patient's Global Assessment of Arthritis

Subjects will complete a global assessment of disease activity using the patient global assessment of disease activity (PtGA) item, a continuous VAS with anchors "0" (very well) to "100" (very poor).

No imputations for missing data will be done.

3.3.6 Physician's Global Assessment of Arthritis

Physicians will complete a global assessment of disease activity using the physician global assessment of disease activity item (PhGA), a continuous VAS with anchors "0" (none) to "100" (extremely active).

No imputations for missing data will be done.

3.3.7 SDAI

The Simple Disease Activity Index (SDAI) is a composite score consisting of the sum of SJC28, TJC28, PtGA/10, PhGA/10, and CRP (mg/dl). Higher values represent higher disease activity.

If one of the components is missing at an individual assessment point, the SDAI value for that assessment will be set to missing.

3.3.8 CDAI

The Clinical Disease Activity Index (CDAI) is a composite score consisting of the sum of SJC28, TJC28, PtGA/10, PhGA/10. CDAI ranges from 0 to 76 with higher values representing higher disease activity. Remission is achieved for a non-missing CDAI value $\leq$2.8.

If one of the components is missing at an individual assessment point, no imputations will be done and the CDAI value for that assessment will be set to missing.

CDAI remission will not be summarized or analyzed but will only be included in analysis datasets and listings. Post-hoc analyses may be performed.

3.3.9 Index- and Boolean-based ACR/EULAR Remission Rates

Boolean-based remission (Felson, 2011) is achieved if all of the following requirements are met at the same time:

- TJC68 $\leq$ 1
- SJC66 $\leq$ 1
- CRP $\leq$ 1 mg/dl
- PtGA $\leq$ 10

If one of the components is missing at an individual assessment point, Boolean-based remission for that assessment will be set to missing.

Index-based remission is achieved if the following requirement is met:

- SDAI $\leq$ 3.3

If the SDAI value is missing at an individual assessment point, Index-based remission for that assessment will be set to missing.

Index- and Boolean-based remission will not be summarized or analyzed but will only be included in analysis datasets and listings. Post-hoc analyses may be performed.

3.3.10 HAQ-DI

The functional status of the subject will be assessed by means of the Disability Index of the Stanford Health Assessment Questionnaire (HAQ-DI). This 20-question instrument assesses the degree of difficulty a person has in accomplishing tasks in eight functional areas:

- dressing & grooming, rising, eating, walking, hygiene, reach, grip, and common daily activities.

Each functional area contains at least two questions. For each question, there is a four level response set that is scored from 0 (without any difficulty) to 3 (unable to do). If aids or devices or physical assistance are used for a specific functional area and the maximum response of this functional area is 0 or 1 the according value is increased to a score of 2.

Table 2 HAQ-DI Aids

Aid or equipment	Will be associated with category score
Walking stick/frame, crutches, wheelchair	Walking
Aids used for dressing	Dressing and grooming
Specially adapted utensils	Eating
Specially adapted chair	Rising
Raised toilet seat, bath rail, bath seat	Hygiene
Long-handled appliance in bathroom	Hygiene
Long-handled appliance for reaching	Reach
Jar opener	Grip
Other (1)	Dressing & grooming, rising, eating, walking
Other (2)	hygiene, reach, grip, common daily activities

If “other” is marked as an aid or equipment then this can be assigned to a group of four functional areas and will be handled as an aid or equipment for each of the four functional areas. Therefore, if the maximum score of a functional area is 0 or 1 that value is increased to a score of 2 for each of the four functional areas.

Regarding these corrections, the highest response within each functional area determines the score of that specific functional area. If no questions within a given functional area were answered, no score will be provided for that category (even if answers on aids or equipment are available).
HAQ-DI is only calculated if there are at least 6 functional area scores available.

The average of these non-missing functional area scores defines the continuous HAQ-DI score ranging from 0 to 3. If there are less than 6 functional area scores available, no imputation will be done and the HAQ-DI will be set to missing for the according assessment.

Severity of pain within the past week is also assessed by the HAQ-DI questionnaire but will not be considered for calculating the HAQ-DI score. Severity of pain will not be included in summaries but will be listed.

3.3.11 Borg Dyspnea Scale

The Borg dyspnea scale will be used to assess dyspnea. A discrete 18-unit scale will be used ranging from "Nothing at all" to "Absolute maximum".

No imputations will be done for missing data.

3.3.12 RAMRIS

All timepoints will be semi-quantitatively reviewed, and the independent reviewer will be responsible for assessing the following features:

Synovitis
For synovitis a total of 8 joints will be evaluated. Individual joint scores range from 0-3. The final synovitis score is the sum of the individual joint scores. If an individual location is scored either 'Not Visible' or 'Surgically Modified' then the score for that location will be set to missing. Missing joint scores will be imputed as the mean of the non-missing joint scores.

Bone Erosion
For bone erosion a total of 25 locations will be evaluated. Individual location scores range from 0-10. The final bone erosion score is the sum of the individual location scores. If an individual location is scored either 'Not Visible' or 'Surgically Modified' or 'Not Assessable' then the score for that location will be set to missing. Missing joint scores will be imputed as the mean of the non-missing location scores.

Bone Edema/Osteitis
For bone edema/osteitis a total of 25 locations will be evaluated. Individual location scores range from 0-3. The final bone edema/osteitis score is the sum of the individual location scores. If an individual location is scored either 'Not Visible' or 'Surgically Modified' or 'Not Assessable' then the score for that location will be set to missing. Missing joint scores will be imputed as the mean of the non-missing location scores.

Cartilage Loss
For cartilage loss a total of 20 locations will be evaluated. Individual location scores range from 0-4 (with increments of 0.5). The final cartilage loss score is the sum of the individual location scores. If an individual location is scored either 'Not Visible' or

'Surgically Modified' then the score for that location will be set to missing. Missing joint scores will be imputed as the mean of the non-missing location scores.

If data for more than 50% of the joints are missing at the time of a given assessment, then the total count will be set to missing for that visit. This threshold might be altered after review of the actual raw data.

Scoring tables of the individual features can be found in Appendix E (Section 6.5). A detailed description of the assessment of RAMRIS parameters can be found in the Independent Review Charter.

3.3.13 RAMRIQ

RAMRIQ measures consist of the quantitative measurements from the following bones and synovial capsules of joints from the affected/dominant hand.

- MCP joints and bones 2-5
- Carpal joints and bones
- Wrist
- Radiocarpal joint (including radial bone)
- Distal radioulnar joint (including ulnar bone)

Different parameters are quantitatively assessed as follows:

Synovitis
Synovitis will be presented as a total measure which is calculated by summing the individual measure of the Volume of Enhancing Pannus (VEP) for the following joints at each time point:

- MCP joints and bones 2-5
- Carpal joints and bones
- Wrist
- Radiocarpal joint (including radial bone)
- Distal radioulnar joint (including ulnar bone)

Synovitis (Normalized)
Normalized synovitis will be calculated using the VEP:

$$\frac{\sum_{all\ joints}(VEP)}{\sum_{all\ joints} Joint\ Volume}$$

The following joints will be considered:

- MCP joints and bones 2-5
- Carpal joints and bones
- Wrist
- Radiocarpal joint (including radial bone)
- Distal radioulnar joint (including ulnar bone)

Erosive Damage (Normalized)
Normalized erosive damage will be calculated as:

$$\frac{\sum_{all\ bones} Volume\ of\ Bone\ Erosion}{\sum_{all\ bones} Bone\ Volume}$$

The extremities of the bones at the following joints will be considered:

- Carpal joints and bones
- Radiocarpal joint (including radial bone)
- Distal radioulnar joint (including ulnar bone)

Cartilage Space Loss
Cartilage space loss will be presented as the sum of the individual measurements (in mm) from the following joints:

- Carpal joints and bones
- Radiocarpal joint (including radial bone)
- Distal radioulnar joint (including ulnar bone)

Bone Marrow Lesions (Normalized)
Bone marrow lesions will be presented normalized by the volume of the bone:

$$\frac{\sum_{all\ bones} Volume\ of\ Bone\ Edema}{\sum_{all\ bones} Bone\ Volume}$$

The extremities of the bones at the following joints will be considered:

- MCP joints and bones 2-5
- Carpal joints and bones
- Radiocarpal joint (including radial bone)
- Distal radioulnar joint (including ulnar bone)

For RAMRIQ, missing joint scores will be imputed as the mean of the non-missing location scores.
If data for more than 50% of the joints are missing at the time of a given assessment, then the total count will be set to missing for that visit. This threshold might be altered after review of the actual raw data.

3.3.14 DCE-MRI

Dynamic Contrast-Enhanced Magnetic Resonance Imaging (DCE-MRI) parameters are measured at each individual joint and consist of:

- Exchange Rate (K^{trans})
- Interstitial Volume (V_e)

- Plasma Volume (V_p)
- Initial Rate of Enhancement (IRE)
- Maximum signal intensity enhancement (ME).

All DCE-MRI parameters will be presented as a total measure over all joints at each time point.
DCI-MRI parameters will be calculated as:

$$\frac{\sum_{all\ joints}(Parameter)}{Number\ of\ Joints}$$

Parameters will also be presented normalized by the VEP.
Normalized DCE-MRI parameters will be calculated as:

$$\frac{\sum_{all\ joints}(VEP\ *Parameter)}{\sum_{all\ joints} VEP}$$

If parameters are missing for individual joints, these joints will be excluded from the numerator and denominator of the calculation.
If data for more than 50% of the joints are missing at the time of a given assessment, then the total count will be set to missing for that visit. This threshold might be altered after review of the actual raw data.

4 STATISTICAL METHODS

4.1 Data Quality Assurance

All tables, figures and data listings to be included in the report will be independently checked for consistency, integrity and in accordance with standard PAREXEL procedures.

4.2 General Presentation Considerations

For each subject 'Baseline' is defined as the last available assessment prior to the start of study treatment. For efficacy variables no Screening information will be used for Baseline. For each subject 'End of Study' is defined as the visit with the last available assessment which is on or after the day of last study treatment.

The relative study day will be included in AE data, medical history and concomitant medication listings, and will be calculated as follows:

- Relative Day 1 is the date of first study medication administration.
- Relative Day of date X = date X - Date of first study medication administration + 1 if date X is on or after date of first study medication administration.
- Relative Day of date X = date X - Date of first study medication administration if date X is before date of first study medication administration.

Relative days before first study medication administration will have the prefix "-". Additionally for relative days after last study medication administration, the number of days since last study medication administration will be presented with the prefix "+". Calculations of "Relative Day" will not include partial dates, but will be left blank in listings.

According to the design the study can be split into 2 time periods:

- Baseline to <Week 12: Day of first study medication administration up to and including the day before Week 12 visit
- Safety Period: Week 12 visit to end of study

Safety Period starts with the day of the Week 12 Visit. If the Week 12 visit is missing, Safety Period starts 14 days after the last study drug administration.

Continuous data will be summarized in terms of the mean, SD, median, minimum, maximum and number of observations, unless otherwise stated. The minimum and maximum will be reported to the same number of decimal places as the raw data recorded in the database. The mean, median, lower quartile and upper quartile will be reported to one more decimal place than the raw data recorded in the database. The SD will be reported to two more decimal places than the raw data recorded in the database. In general, the maximum number of decimal places reported shall be four for any summary statistic.

Categorical data will be summarized in terms of the number of subjects providing data at the relevant time point (n), frequency counts and percentages.

If not stated otherwise percentages will be presented with no decimal place. Percentages will not be presented for zero counts. Percentages <1 will be shown as '<1'. Percentages >99 but <100 will be shown as '>99'. Percentages will be calculated using n as the denominator.

P-values greater than or equal to 0.001, in general, will be presented to three decimal places. P-values less than 0.001 will be presented as "<0.001".

Confidence intervals (CI) will be presented to one more decimal place than the raw data.

All report outputs will be produced using SAS® version 9.2 or a later version in a secure and validated environment. Tables, listings and figures will each be provided in separate rtf documents as well as in combined pdf documents.

4.3 Study Subjects

4.3.1 Disposition of Subjects

A clear accounting of the disposition of all subjects who enter the study will be provided, from screening to study completion.

The following subject disposition summaries will be provided: See Section 4.5 for a description of the analysis populations.

A summary of the number of subjects who were screened for entry into the study and

- were excluded prior to randomization by major reason and overall (Analysis population: All Subjects Screened)
- were randomized per center, per country, and by disease duration (≤2 years, >2 years) and treatment group (Analysis population: Randomized Population)
- received at least one dose of study medication by treatment group (Analysis population: Randomized Population)
- withdrew from the study by major reason of withdrawal and by treatment group (Analysis population: Randomized Population)
- completed 12 weeks of study treatment by treatment group (Analysis population: ITT Population)
- completed the study by treatment group (Analysis population: ITT Population)

By-subject listings of eligibility details, randomization details by site (including subject id, randomization number and treatment group and information whether the blind was broken at discontinuation), visit dates (including actual treatment received at each visit) and withdrawal/study completion details (including reason for discontinuation and duration of treatment prior to discontinuation) will be provided.

4.3.2 Protocol Deviations

Protocol deviations will be managed and reported according to PAREXEL Standard Operating Procedures (SOPs).

Details will be given in the Protocol Deviation Specification (PDS) document, which will include detailed information regarding definitions, classifications, tracking and management of PDs.

PDs will be categorized as major or minor as further described in the PDS.

PDs will also be classified into the following categories: Informed Consent, Inclusion/Exclusion Criteria, Withdrawal Criteria, IP Admin/Study Treat, Disallowed Medications, AE/SAE, Visit Schedule, Procedures/Tests, Other.

Observable PDs that can only be identified by project team members during the monitoring, but cannot be programmed, will be tracked and reconciled with programmable (data driven) PDs, that can be programmed from the recorded data.

A combined PD dataset will include all PDs.

Special considerations are needed for (potentially) unblinding PDs, which will be handled in separate unblinded areas with restricted access.

Summaries of the number and percentage of randomized subjects with a major/minor PD will be provided by treatment group. Subjects with a major PD will be summarized by type of deviation.

A by-subject listing of all protocol deviations will be provided.

4.4 General Data Handling Conventions

4.4.1 Missing Dates

In analysis of AEs and medication a complete date will be established in order to identify AEs or medication as occurring during treatment or not. For handling partially reported onset/start and outcome/end dates for AEs or medication the following algorithms are applied:

- AEs:
  - Missing onset day, but month and year present:
    If study medication had been taken in the same month and year as the occurrence of the AE, then the onset day of the event is assigned to the day of first dose of study medication.
    Otherwise the onset day is set to the first day of the month (e.g., XX-Sep-2010 is considered as 01-Sep-2010).
  - Missing onset day and month, but year present:
    If study medication had been taken in the same year as the occurrence of the AE, then the onset date of the event is assigned to the date of first dose of study medication.
    Otherwise the onset day and month is set to 01 January (e.g., XX-XXX-2010 is considered as 01-Jan-2010).
  - Missing outcome day, but month and year present:
    The day is set to the last day of the month (e.g., XX-Sep-2010 is considered as 30-Sep-2010).
  - Missing outcome day and month, but year present:
    The outcome day and month is set to 31 December (e.g., XX-XXX-2010 is considered as 31-Dec-2010).
- Medications:
  - Missing start day, but month and year present:
    If first study medication administration had been taken place in the same month and year as the occurrence of the medication, then the start day of the medication is assigned to the day of first study medication administration. Otherwise the start day is set to the first day of the month (e.g., XX-Sep-2010 is considered as 01-Sep-2010).
  - Missing start day and month, but year present:

If first study medication administration had been taken in the same year as the occurrence of the medication, then the start date of the medication is assigned to the date of first study medication administration.
- Otherwise the start day and month is set to 01 January (e.g., XX-XXX-2010 is considered as 01-Jan-2010).
- Missing stop day, but month and year present:
  The day is set to the last day of the month (e.g., XX-Sep-2010 is considered as 30-Sep-2010).
- Missing stop day and month, but year present:
  The stop day and month is set to 31 December (e.g., XX-XXX-2010 is considered as 31-Dec-2010).

- RA diagnosis and symptom onset:
  - Missing day, but month and year present:
    Date is set to the first day of the month (e.g., XX-Sep-2010 is considered as 01-Sep-2010).
  - Missing day and month, but year present:
    Date is set to 01 January (e.g., XX-XXX-2010 is considered as 01-Jan-2010)
- End date of Tobacco use:
  - Missing outcome day, but month and year present:
    The day is set to the last day of the month (e.g., XX-Sep-2010 is considered as 30-Sep-2010).
  - Missing outcome day and month, but year present:
    The outcome day and month is set to 31 December (e.g., XX-XXX-2010 is considered as 31-Dec-2010).

In data listings, onset/start and outcome/stop date of AEs, medication, diagnosis, or symptoms is displayed as reported.

4.4.2 Missing Data for Efficacy and Safety Endpoints

4.4.2.1 Missing Data for Efficacy Endpoints

Missing data for each efficacy endpoint will be handled as described in Section 3.3. Section 4.5.1 describes another rule for handling missing efficacy data for statistical and summary analyses.
For each continuous endpoint, change from baseline will be missing at visits with missing post-baseline values or where data were imputed to missing.

For efficacy endpoints (including CRP and ESR), a missing value at baseline will not be imputed with the screening results for change from baseline or percent change from baseline calculations; the endpoint will be missing for all visits.

4.4.2.2 Missing Data at Baseline for Safety Endpoints

For safety endpoints, a missing value at baseline will be imputed using the prior result that is closest in proximity to baseline in order to calculate change from baseline or percent change from baseline. No imputation will be done for post-baseline visits.

4.4.3 Multiple Assessments and Early Withdrawal Visits

If a variable has been assessed multiple times at the same visit, only the last assessment will be used except for Electrocardiograms (ECGs) or laboratory measurements. Triplicate ECG measures will be averaged for each subject and visit prior to generating summary tables.

For laboratory values the value with the worst Common Terminology Criteria for Adverse Events v 4.0 (CTCAE) grade will be considered. If there is more than one value of this grade, the later value will be considered.

Only scheduled visits will be included in summaries. Listings will include scheduled and unscheduled visits.

Early Withdrawal (EW) visits will be assigned to the respective visit according to the Time and Events table in Section 3.1.2 using the day of the EW visit relative to Day 1. For the first 4 weeks an EW visit more than ±3 days in comparison to the target visit will be assigned to the closest target visit. Starting with Week 4, an EW visit more than 7 days after a target visit day will be assigned to the next visit.
If for a certain variable a scheduled visit and an EW visit occur within the same visit window and are both available, the later visit will be used for analyses. The earlier visit will appear in listing only. For efficacy analyses, if an EW visit is assigned to a non-standard efficacy visit, i.e., a visit at which efficacy assessments are not scheduled per the Time and Events Table, then these EW data will be ignored in the statistical analysis.

4.4.4 Visit Windows for Efficacy Parameters

The following rules will be implemented for clinical efficacy parameters:

- CRP/ESR must be pre-dose for each visit (day of dosing is assumed to be pre-dose, if time of assessment unavailable).
- Baseline must be pre-dose (day of dosing is assumed to be pre-dose, if time of assessment unavailable).
- Assessments on consecutive days can be included in the calculation of the composite score.
- Assessments more than 1 day apart should not be used in the same composite endpoint. Parameters that violate this rule will not be considered for derivations of composite efficacy scores (e.g. DAS28, ACR).

For the Week 4 and Week 12 visits, the MRI assessment should not be performed more than 7 days after the day of the respective visit. The day of the visit will be defined as the

day of study drug administration or (if no study drug was administered) the day when the majority of the components of the DAS28( ) score were assessed. Assessments that violate this rule will not be considered for summaries or analyses (i.e. Tables or Figures).

4.4.5 Oral Corticosteroid Conversion

To assess corticosteroid use and determine average daily corticosteroid dose, all corticosteroid dosages will be converted to a prednisone equivalent in milligrams by multiplying the dose of the steroid (using the coded term from GSKDrug) by the conversion factor to get prednisone equivalent units. Summaries refer to average daily prednisone dose instead of average daily corticosteroid dose. See Table 3 for conversion factors.

Table 3 Corticosteroid Conversion Factors

Conversion	Steroid
8.333	BETAMETHASONE
8.333	BETROSPAM
0.2	CORTISONE
8.333	CRONOLEVEL
0.83333	DEFLAZACORT
6.667	DEXAMETHASONE
0.25	HYDROCORTISONE
1.25	MEPREDNISONE
1.25	METHYLPREDNISOLONE
2.5	PARAMETHASONE
1	PREDNISOLONE
1	PREDNISONE
1.25	TRIAMCINOLONE

For each scheduled visit, the average daily prednisone dose will be calculated by summing all oral corticosteroid doses (converted to prednisone) since the previous visit up to and including the current visit and then dividing by the number of days in this period (Date of current visit – Date of previous scheduled visit).

4.5 Analysis Populations

Screened Population: The Screened Population will consist of all subjects who were screened.

Randomized Population: The Randomized Population will consist of all subjects who were randomized to treatment.

Intent-to-Treat (ITT) Population: The ITT population will consist of all subjects who were randomized to treatment and who received at least one dose of study treatment (GSK3196165 or placebo).

Pharmacokinetic (PK) Population: The PK population will consist of all subjects in the ITT population, who had at least one valid PK assessment.

Immunogenicity Population: The immunogenicity population will consist of all subjects in the ITT population, who had at least one valid immunogenicity assessment.

Genetic Research Population (GRP): The GRP will consist of all subjects in the ITT population who consent to participate in the genetic research. Any subject who has received an allogeneic bone marrow transplant will be excluded from the genetic research population.

All endpoints will be analysed for the ITT population by treatment group. Numbers and percentages of subjects in each analysis population will be summarized by treatment group. Subjects who received at least one wrong dose will be presented in a listing. Additional summaries may be provided if considered necessary.

A by-subject listing of analysis population details will be provided. This listing will be presented by treatment group and will include: center, subject identifier, inclusion/exclusion flag for each population and reason for exclusion from each population. All subjects screened will appear on this listing.

4.5.1 Analysis Datasets

For each of the clinical efficacy evaluations of binary endpoints, subjects with missing efficacy data, early withdrawals, subjects who received more than 10 mg/day of prednisone equivalent oral corticosteroids, subjects with any new use of parenteral corticosteroids or with an intra-articular corticosteroid injection will be imputed as non-responders and therefore treated as a failure (e.g. no DAS28(CRP) remission or no moderate/good EULAR response).

A dataset will be created with these imputation rules and will be referred to as the NRI dataset. The NRI dataset will be the primary dataset used for all binary efficacy endpoints. Imputed response for each binary endpoint will be presented by visit and subject in listings.

For the clinical efficacy evaluations of continuous endpoints, results recorded for subjects who exceeded 10 mg/day prednisone equivalent oral corticosteroids, subjects with any new use of parenteral corticosteroids or with an intra-articular corticosteroid injection will be set to missing at the same visits that were set to non-responder for the binary endpoints. A dataset will be created and will be referred to as the Set-to-Missing (STM) dataset.

Both the STM and the NRI dataset will also contain observed data.

The following imputation rules will be followed for STM and NRI datasets:

- Missing clinical efficacy data: visits with missing data (due to a missed visit or missing component of a composite endpoint) should be set to treatment failure.
- Early withdrawals: only visits following the early withdrawal visit should be set to treatment failure.
- Prednisone equivalent >10 mg/day: the visit after a period (time from previous visit up to the day before the current visit) of average daily corticosteroid use exceeding 10 mg/day prednisone and all following visits should be set to treatment failure (continuous endpoints will be set to missing).
- Parenteral corticosteroids: all visits after a subject received any new parenteral corticosteroids (e.g. intravenous and intra-muscular) should be set to treatment failure (continuous endpoints will be set to missing).
- Intra-articular corticosteroids: all visits after an intra-articular corticosteroid injection should be set to treatment failure (continuous endpoints will be set to missing).

4.6 Demographic and Other Baseline Characteristics

The following Baseline characteristics will be summarized based on the ITT population by treatment group and overall.

Demographic Characteristics
Continuous variables: age (years), height (cm), weight (kg), BMI-Body Mass Index (kg/m²).

Categorical variables: age (<18 years, 18- <65 years and ≥65 years), sex (male, female), child bearing potential (for females: yes, no), ethnicity (Hispanic or Latino, Not Hispanic or Latino), race (African American/African Heritage, American Indian or Alaskan Native, Asian – Central/South Asian Heritage, Asian – East Asian Heritage, Asian – Japanese Heritage, Asian – South East Asian Heritage, Native Hawaiian or Other Pacific Islander, White – Arabic/North African Heritage, White – White/Caucasian/European Heritage) and BMI (<18.5, 18.5 - <25, 25 - <30, ≥30).

Baseline efficacy parameters
Continuous baseline efficacy parameters and components will be summarized in a separate table.

Medical history
Prior and concomitant medical history will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) in the latest available version. Summary tables will be presented by system organ class (SOC) and preferred term (PT).

Disease history
RA disease history will be summarized in form of disease duration categorical (<≤ 2 years/ >2 years) and as summary statistics, RA functional class and time since start of RA symptoms. Disease duration will be calculated as duration from formal RA diagnosis

to first study treatment administration. For subjects who got randomized but never received study treatment, disease duration will be calculated up to the date of randomization.

Separate by-subject listings of demographic data and other baseline characteristics, prior and concomitant medication, medical history, RA medication use and disease history will be provided.

Rheumatoid Factor
Baseline rheumatoid factor (RF) will be summarized as numbers and percentages for positive and negative.

ACPA Status
Baseline ACPA status will be summarized as numbers and percentages for positive and negative.

4.7 Medication Use

Prior and concomitant medication use
Prior and concomitant medications will be summarized by Anatomical Therapeutic Chemical (ATC) classification, ingredient and treatment group in separate tables. Medications that are both prior and concomitant will appear in both tables.

Medications that start and stop prior to the date of first dose of study medication will be classified as Prior only. If a medication starts before the date of first dose of study medication and stops on or after the date of first dose of study medication, then the medication will be classified as both Prior and Concomitant. Medications will be classified as Concomitant only if they have a start date on or after the date of first dose of study medication.

If medication start and/or stop dates are missing or partial, imputations are done as described in Section 4.4.1 and the dates will be compared as far as possible with the date of first dose of study medication. Medications will be assumed to be Concomitant only, unless there is clear evidence (through partial dates as described in Section 4.4.1 to suggest that the medication started prior to the first dose of study medication. If there is clear evidence to suggest that the medication started prior to the first dose of study medication, the medication will be assumed to be both Prior and Concomitant, unless there is clear evidence to suggest that the medication stopped prior to the first dose of study medication. If there is clear evidence to suggest that the medication stopped prior to the first dose of study medication, the medication will be assumed to be Prior only.

Prior and concomitant corticosteroid use will be shown in a separate listing.

RA medication use
Prior and concomitant RA medication will be summarized separately by ATC classification, ingredient and treatment group.

Corticosteroid use
Change from baseline in average daily prednisone dose will be summarized by visit. The number and percentage of subjects with ≥25% decrease from baseline and to below ≤7.5 mg/day will be summarized by visit.
Subjects who received corticosteroids will be summarized by treatment group and route of administration.

4.7.1 Extent of Exposure

The extent of exposure to GSK3196165 will be evaluated by summarizing the number of injections, the volume of doses received in ml and the duration of drug exposure in days. Duration of drug exposure will be calculated as:

- date of the last injection during study – date of first injection during study + 14 days

Weekly MTX dose will be averaged across the entire study for each subject.
MTX dose should be documented in mg, but in case ml is captured the following conversion should be used: 1ml is equivalent to 25 mg.

Extent of exposure for study medication and MTX will also be presented in by-subject listings.

4.8 Treatment Compliance

Summaries will be based on the ITT population and will be provided by treatment group.

Study Treatment
Treatment compliance for GSK3196165 and Placebo will utilize the administered volume and compare it to the scheduled expected volume. The general formula for the compliance ratio (CR) is given as follows:

$$\mathrm{CR} = \frac{\text{actual volume (mL)} * 100}{\text{expected volume (mL)}}\%$$

The expected volume is based on the planned number of doses (based on each subject's time in the study) and the study treatment volume per dose (1.2mL). If a subject completed the study, then the total number of doses would be 8 and the total expected volume for each treatment group would be 9.6mL each for Placebo and GSK3196165 180 mg. The actual volume is the sum of the volume of doses received in mL. The ratio will be summarized as a continuous variable as well as categorically (<80% and ≥80%). Visits after an early withdrawal or an AE-caused study drug discontinuation will not be included in the denominator.

Additionally, number and percentage of subjects who missed at least two doses of study treatment will be summarized.

Methotrexate
Numbers and percentages of subjects who increased or decreased their MTX dose or missed a MTX dose during the conduct of the study will be summarized.

A by-subject listing of treatment compliance ratio for GSK3196165 will be provided.

4.9 Efficacy Evaluation

4.9.1 Analysis and Data Conventions

4.9.1.1 Multi-center Studies
This study will randomize approximately 40 subjects into 2 treatment arms, with approximately 30 subjects in GSK3196165 treatment group and approximately 10 subjects in placebo group. The number of sites (centres) may be roughly around 15, so that an adjustment of the analyses by centre will not be feasible.

For all analyses all sites, countries and regions will be pooled.

4.9.1.2 Adjustments for Covariates
Disease duration (early/established RA) will be considered in all efficacy analyses.

4.9.1.3 Multiple Comparisons/Multiplicity
All endpoints are exploratory and no interim analyses are planned. Hence no adjustments for multiplicity are required.

4.9.1.4 Interim Analyses
No interim analyses are planned for this study.

4.9.1.5 Examination of Subgroups
Subgroup analyses will not be done as part of the primary analysis but may be considered for post hoc analyses.

The following subgroups will be implemented in analysis datasets for clinical efficacy and biomarker endpoints.:

- early RA (i.e. ≤ 2 years disease duration before first study medication administration)
- baseline rheumatoid factor (positive/negative)
- baseline ACPA status (positive/negative)

4.9.2 Primary Efficacy Endpoint – Exploratory Biomarkers

Change from baseline in exploratory biomarkers.

Summary statistics and plots of change over time for the absolute values and change from baseline will be created for all exploratory biomarkers. These will be grouped by type of biomarker (see section 6.1 in Appendix A):

- Target Engagement
- Predictive Biomarkers
- Complement Biomarkers
- Cartilage Biomarkers
- Mechanistic Biomarkers
- Flow Cytometry: 6 Colour TBNK Panel
- Flow Cytometry: CD16+ Monocyte Panel
- Flow Cytometry: T Reg Cell Foxp3
- T Helper Cell Panel
- Safety Biomarkers

For selected biomarkers (see Table 5), a repeated measures model analysis of the observed value and change from baseline for visits at Weeks 1, 2, 4, 6, 8, 12 and 22 (12-week follow-up) will be conducted for the difference between treatments. The model will be fitted with fixed effects for randomised treatment group, disease duration (≤2 years or >2 years), visit, treatment by visit interaction and baseline value, and visit within subject as a repeated effect. Observed margin based on the population that contributes to the analysis for each biomarker will be used. This requires the OM dataset to contain one record per subject per visit with all covariates merged onto each record. The covariance will initially be specified as unstructured covariance matrix. If SAS gives a non-convergence warning, the results will not be used – instead the warning will be stored and the spatial power law structure will be used in order to reduce the number of parameters to be estimated.

The point estimates and corresponding 95% CI for the treatment differences will be constructed using the residual error from the model; these estimates will be summarised and plotted for each randomised treatment group comparison over time.

Relationship of exploratory biomarker to clinical endpoints

The relationship between selected exploratory biomarkers and the following clinical endpoints will be explored:

- ACR50 at Week 12
  - If too few subjects meet ACR50 criteria, then ACR20 may be used in replacement.
- DAS28(CRP) score up to Week 22
- Pain score up to Week 22

Summary statistics of absolute value and change from baseline for all biomarkers will be presented separately by ACR50 response at Week 12. Box plots of biomarker response over time, grouped by binary clinical response and treatment will be presented.

A scatter plot matrix of continuous endpoints (DAS28(CRP) and pain score) against biomarker will be plotted and will include a simple linear regression line to aid with interpretation. All plots will be presented for overall and by treatment.

Relationship Between Biomarkers

The relationship between % TH17 (CD45+CD3+CD8-CD4+CCR6-CXCR3-) and the following parameters will be plotted on a scatter plot matrix.

- T regulatory cells (%CD3+CD4+CD25+CD127-FOXP3+)
- IL-17A
- IL-17F

Post-hoc analyses

For biomarkers of further interest exploratory analysis using an appropriate statistical model to identify any trends may be carried out in post-hoc analyses. The model will determine whether the mechanistic / biomarker effect significantly explains or predicts the effect on the clinical endpoints. This may be conducted through comparing statistical models; incorporating different explanatory terms (i.e., mechanistic / biomarker endpoints) with the 'null' model (no mechanistic / biomarker endpoints); or, if deemed appropriate, multivariate statistical methods may also be applied to determine the relationship between the key endpoints. The consistency in the changes over time between the endpoints will also be assessed.

Log transformation

Each of the biomarkers will be examined to determine if a log transformation should be applied prior to any summary or statistical analyses. If required, then Geometric means and CV% will be presented in summary statistics and change from baseline will be presented as a ratio. Either original scale, Log2 or Log10 will be used for graphical presentation whichever is deemed more appropriate for interpretation.

4.9.3 Secondary Efficacy Endpoints

4.9.3.1 Inflammatory Structural Joint Damage in the Hand/Wrist

Absolute and change from baseline in RAMRIS synovitis will be analyzed with a Bayesian approach. Change from baseline will be analyzed for the difference between treatments over visits at Week 4, 12 and 22 (12-week follow-up). The model will be fitted with fixed effects for disease duration (<=2 years or >2 years), treatment by visit interaction, baseline value, and visit within subject as a repeated effect. The baseline value will be centered and an observed margin based on the population that contributes to analyses will be used for the disease duration. This requires the OM dataset to contain one record per subject per visit with all covariates merged onto each record. A non-informative prior will be used. Details of the methodology are shown in Appendix C (Section 6.3).

Additionally, a second approach will be used for synovitis in RAMRIS. A historical data prior will be applied for placebo and change from baseline at Week 12. A Bayesian

model for the change from baseline at Week 12 will be fitted with fixed effects for disease duration (<=2 years or >2 years), treatment, and baseline value. The baseline value will be centered and an observed margin based on the population that contributes to analyses will be used for the disease duration. This requires the OM dataset to contain one record per subject per visit with all covariates merged onto each record.

Historical Data Prior

Placebo + MTX data from 3 trials was used to inform the historical placebo response: GO-FORWARD [Conaghan (2011)], GO-BEFORE [Østergaard (2011)] and RA-SCORE [Peterfy (2016)]. The data was pooled using a random effects meta analyses and assumed to be equivalent of one study with 60 subjects.

Prior (Placebo) ~ Normal (-0.116, SD=1.463)

The prior will be down-weighted in the Bayesian analyses using a power-prior approach with a weight of 50%.

4.9.3.2 Exploratory MRI/Imaging Endpoints

RAMRIS bone erosion, osteitis, cartilage loss, and RAMRIQ synovitis, synovitis normalized, erosive damage normalized, cartilage space loss, and bone marrow lesions normalized will be analyzed with a repeated measures model as described for the clinical efficacy parameters in Section 4.9.3.3.

Absolute and change from baseline in joint inflammation as measured by DCE-MRI in the hand/wrist:

- K^{trans}
- V_e
- V_p
- IRE
- ME

The DCE-MRI parameters will be analyzed with a repeated measures model as described for the clinical efficacy parameters in Section 4.9.3.3.

4.9.3.3 Clinical Efficacy Variables

All clinical efficacy outputs will be created using the ITT population with NRI or STM rules as described in 4.5.1.

The following binary endpoints will be presented by numbers and percentages for all assessment points:

- ACR20, ACR50 and ACR70 response
- DAS28(CRP) remission
- EULAR response (binary: good/moderate, no response)

- DAS28(CRP) LDA

For each of the categorical endpoints the difference (%) to placebo, odds ratio's and confidence intervals will be presented. P-values will not be presented. Odd's ratios calculated from a logistic regression model will include terms for treatment group, baseline and disease duration (<=2 years or >2 years).

The following continuous endpoints will be summarized and analyzed as observed values and change from baseline for all assessment points:

- DAS28(CRP)
- TJC28, SJC28, TJC68, SJC66
- CRP value
- Patient's Global Assessment of Arthritis Disease Activity (PtGA)
- Patient's Assessment of Arthritis Pain
- Physician's Global Assessment of Arthritis (PhGA)
- SDAI
- CDAI
- HAQ-DI

Plots of Mean and 95% CI of Change from Baseline in DAS28(CRP) at Each Visit will be created.
Subject profile plots of DAS28 score at each visit will be produced.

For all continuous endpoints, a mixed model repeated measures (MMRM) analysis will be conducted including all weeks through Week 22. The model will be fitted with fixed effects for disease duration (≤2 years or >2 years), randomised treatment group, visit, treatment by visit interaction and baseline, and visit within subject as a repeated effect. The covariance will initially be specified as unstructured covariance matrix. Observed margin based on the population that contributes to analyses will be used. This requires the OM dataset to contain one record per subject per visit with all covariates merged onto each record. If SAS gives a non-convergence warning, the results will not be used – instead the warning will be stored and the spatial power law structure will be used in order to reduce the number of parameters to be estimated.

The point estimates and corresponding 95% CI for the treatment differences will be constructed using the residual error from the model; these estimates will be summarized and plotted for each randomised treatment group comparison over time.

4.9.4 Overview of Summaries and Analyses of Efficacy Endpoints

Tables 4 and 5 show the outputs that will be provided for summaries and analyses of the efficacy endpoints.

Table 4 Summaries of Binary Efficacy Endpoints

	Summary Statistics	
	Frequencies/ OR/ CI*	Proportions
	T	**F**
Clinical		
DAS28(CRP) Remission	Y	Y
DAS28(CRP) LDA	Y	Y
ACR Response	Y	Y
Moderate/Good EULAR Response	Y	Y

CfB=Change from baseline; F=figure; OR=Odds Ratio; T=table
*Wald CIs for difference (%) to placebo and for odds ratio

Table 5 Summaries and Analyses of Continuous Efficacy Endpoints

	Summary: O + CfB		MMRM Analysis		Bayesian MMRM Analysis		by Clinical Parameters*			
							Summary: O + CfB		MMRM Analysis	
	T	**F**	**T**	**F**	**T**	**F**	**T**	**F**	**T**	**F**
MRI										
Inflammatory Structural Joint Damage	Y	Y			Y	Y				
Joint Inflammation	Y	Y	Y	Y						
Clinical										
DAS28(CRP)	Y	Y	Y	Y						
CRP	Y	Y	Y	Y						
HAQ-DI Score	Y	Y	Y	Y						
Pain Score	Y	Y	Y	Y						
TJC28 and TJC68	Y	Y	Y	Y						
SJC28 and SJC66	Y	Y	Y	Y						
PtGA	Y	Y	Y	Y						
PhGA	Y	Y	Y	Y						
Biomarkers										
Target Engagement Biomarkers	Y	Y	Y	Y			Y	Y	Y	Y
Predictive Biomarkers	Y	Y	Y	Y			Y	Y	Y	Y
Complement Biomarkers	Y	Y								
Cartilage Biomarker	Y	Y	Y	Y			Y	Y	Y	Y
Mechanistic Biomarkers	Y	Y								
Flow Cytometry: 6 Colour TBNK Panel	Y	Y	Y	Y			Y	Y	Y	Y
Flow Cytometry: CD16+ Monocyte Panel	Y	Y	Y	Y			Y	Y	Y	Y
Flow Cytometry: T Reg Cell Foxp3	Y	Y	Y	Y			Y	Y	Y	Y
T Helper Cell Panel Foxp3	Y	Y	Y	Y			Y	Y	Y	Y
Safety Biomarkers	Y	Y								

O=Observed, CfB=Change from baseline
* Clinical Parameters are ACR50, DAS28 and Pain Score

4.10 Safety Evaluation

All safety summaries and analyses will be based upon the ITT population and by actual treatment group at the time of a safety event or assessment. Actual treatment group is defined by the last dose a subject received prior to the relevant event/assessment.

4.10.1 Adverse Events

Adverse events will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) in the latest available version.

AESI	Programmatical Derivation
Serious infections, including serious respiratory infections.	SAEs, Filter on infections SOC
Opportunistic infections including TB reactivation	Opportunistic infections will be adjudicated by the Safety Review Team (SRT), using the preferred terms list given in Appendix D. Final adjudication conducted by SRT.
Neutropenia	Based on Grade 3 or 4 absolute neutrophil count
PAP (Pulmonary alveolar proteinosis)	If identified, then list all AEs for those subjects.
Hypersensitivity reactions, including anaphylaxis	Hypersensitivity reactions will be adjudicated by the SRT, using AE data and data from hypersensitivity reactions eCRF page. Final adjudication conducted by SRT.
Injection site reactions	Using data from Injection site reactions eCRF page

AEs with missing intensity will be considered severe. AEs with missing relationship to study medication will be considered as related both to GSK3196165/Placebo and to MTX.

Adverse events will be summarized by system organ class (SOC) and preferred term (PT) Summaries will provide the number of subjects reporting at least one AE and the total number of events reported. AESIs will not be presented by SOC or PT, but by numbers and percentages of the six defined terms for AESIs.

Adverse event summaries will be ordered in terms of decreasing frequency for SOC, and PT within SOC, in the GSK3196165 treatment group, and then similarly by decreasing frequency in the placebo treatment group, and then alphabetically for SOC, and PT within SOC.

The following summaries of AEs will be provided by randomized treatment group:

- Overall summary of AEs, including number and percentages of subjects with
  - Any AE
  - Serious AE
  - Discontinuation due to AE
  - Drug-related AE
  - AE leading to death
- Overall summary of AEs
- Number and percentage of subjects reporting an AE
- Most common AEs (reported by >5% of the subjects in total)
- Plot of most common AEs and Relative Risk
- Number and percentage of subjects reporting an AE by maximum intensity
- Number and percentage of subjects reporting an AE by relationship to study treatment
- Number and percentage of subjects reporting an AESI
- Details of Injection Site Reactions
- Details of Systemic Hypersensitivity Reactions
- Number and percentage of subjects that withdraw from study or discontinue study treatment as result of an AE
- Exposure-adjusted (patient years) AEs by treatment group, where the total number of subjects for each column includes all subjects who took the dose at least once during the study. The exposure-adjusted rate divides the total number of AEs attributed to each dose by duration of exposure (in years) to that dose. The duration of exposure is calculated by (date of last dose – date of first dose + 14).
- AEs by period of first onset
- Serious AEs by period of first onset

Exposure adjusted incidence rates are calculated as X/T*100 with

- X being the number of occurrences of the relevant AE for subjects while being in the corresponding treatment group
- T being the cumulative duration of exposure (in years) for subjects in the corresponding treatment group

Adverse event summaries will be ordered by decreasing frequency for SOC, and PT within SOC, summed up for all subjects and then alphabetically for SOC, and PT within SOC.

A by-subject listing of all AEs will be provided as well as a by-subject listing of all AEs leading to withdrawal from study. These listings will be presented by treatment group and will include: center, subject identifier, age, sex, race, weight, AE (SOC, PT, and verbatim term), date of onset, date of resolution, duration, intensity, action taken, outcome, relationship to study medication, action taken with study medication, other action taken and whether the AE is serious or an AE of interest.

4.10.2 Disease Related Events

The number of disease related Events (DREs) will be summarized by treatment group. A by-subject listing of all DREs that occurred during the study will also be provided.

4.10.3 Respiratory Events

Numbers and percentages of subjects experiencing persistent cough, or persistent dyspnea or persistent D_{LCO} decrease will be reported in a table by randomized treatment group.

Event	Definition
Persistent cough	Cough grade 2 or greater recorded for 3 consecutive weeks (15 or more days) on the eCRF page.
Persistent dyspnea	Borg Scale grade 3 or greater recorded for 3 consecutive weeks (15 or more days) on the eCRF page.
Persistent decrease of D_{LCO} by >15%	Relative decrease of D_{LCO} Hemoglobin-corrected Percent of Predicted of ≥15% compared to Baseline for 3 consecutive weeks (15 or more days)

These definitions are provided for guidance to identify potential cases. Final adjudication will be conducted by the SRT.

4.10.4 Deaths and Serious Adverse Events

The following summaries will be provided:

- Number and percentage of subjects reporting a serious AE by treatment group, SOC and PT
- Number and percentage of subjects reporting a drug related serious AE by treatment group, SOC and PT
- Number and percentage of AEs leading to death

The following listings will be provided:

- A by-subject listing of all Serious AEs
- A by-subject listing of all AEs leading to discontinuation of study treatment
- A by-subject listing of all Serious AEs leading to death

4.10.5 Pulmonary assessments

All pulmonary assessment results (chest X-ray, cough, Borg dyspnea questionnaire, lung auscultation, pulmonary function tests (PFTs - spirometry, gas transfer [D_{LCO}]) and pulse oximetry) will be provided in listings.

Summary statistics will be provided for the change from baseline for the Borg dyspnea score, PFTs, spirometry and D_{LCO}, and for pulse oximetry by visit. The number and percentages of subjects having a cough for each grade will be summarized by visit, in addition to the number and percentages of subjects having a relative decrease in D_{LCO} Hemoglobin-corrected Percent of Predicted (DLCOHCPP) of >15% from screening.

For summary tables, reporting of subjects with >15% decrease from baseline will be shown separately for subjects with current D_{LCO} <70%. A separate summary will be provided by baseline D_{LCO} value (<70% and ≥70%).

Individual patient profiles of D_{LCO} over time will be shown in a figure by treatment group.

Spirometry parameters Forced Vital Capacity (FVC) and Forced Expiratory Volume in 1 Second (FEV1) will be presented in percent of predicted. If FVC is assessed in Litres rather than percent predicted, it will be transformed to percent predicted. The predicted value for each subject will be derived using age, height, sex, race, and ethnicity and reference values as given in [Kankinson (1999)].

4.10.6 Cardiovascular Events

Occurrences of the below cardiovascular events will be listed in 9 separate listings:

- Arrhythmias
- Congestive heart failure
- Cerebrovascular event stroke and transient ischemic attack
- Deep vein thrombosis/ Pulmonary embolism
- Myocardial Infarction / Unstable Angina
- Peripheral arterial thromboembolism
- Pulmonary Hypertension
- Revascularisation
- Valvulopathy

All relevant data captured in the CRF will be listed. However, as only very few events might occur during the conduct of the study, the design of the listings might be adapted to the actual observed data.

4.10.7 Liver Events

Liver chemistry stopping events as defined in the protocol trigger further assessments (e.g. liver PK sampling, assessment of alcohol intake, liver biopsy).

All relevant data captured in the CRF will be listed. However, as only very few events might occur during the conduct of the study, the design of the listings might be adapted to the actual observed data.

4.10.8 Clinical Laboratory Evaluation

The central laboratory will analyse and assess blood and urine samples for the following:

Table 6 Laboratory Assessments

Hematology	Biochemistry	Urinalysis
Hemoglobin	Sodium	**Urine dipstick**
Hematocrit	Potassium	Glucose
Mean cell volume (MCV)	Calcium	Protein
Mean corpuscular hemoglobin (MCH)	Phosphate	
Mean corpuscular hemoglobin concentration (MCHC)	Urea	Microscopy of urine sediment for erythrocytes, leukocytes and casts if urine dipstick abnormal
Erythrocyte count	Creatinine	
Reticulocyte count	Creatinine clearance (calculated)	
Leukocyte count	AST	
Leukocyte differential count	ALT	
neutrophils	γ-glutamyl transpeptidase (GGT)	Urine pregnancy test
eosinophils	LDH	
basophils	Alkaline phosphatase (AP)	
monocytes	Bilirubin (total)	
lymphocytes	Creatine Phosphokinase (CPK)	
Platelets	Total protein	
Activated partial thromboplastin time (aPTT)	Albumin	
Prothrombin Time (PT)	Albumin/globulin ratio	
International Normalized Ratio (INR)	Serum Glucose	
Fibrinogen	CRP	
ESR*	Cholesterol**	
	Triglycerides**	
	HDL**	
	LDL**	

*Measured locally
**Fasting tests

Descriptive statistics for observed values and change from baseline will be presented by treatment for each visit and for all parameters for hematology, biochemistry and urinalysis, separately.

Laboratory values reported below the lower limit of quantification (LLQ) will be replaced by one-half the LLQ when reporting summary statistics. Laboratory values

reported above the upper limit of quantification (ULQ) will be replaced by the ULQ when reporting summary statistics.

Laboratory parameters of interest will be summarized (number and percentages) by time point and by grade as defined in Appendix B. Additionally, a summary of shift from baseline to worst grade during the study value will be provided.

Boxplots will be produced to display the distribution of results for laboratory parameters of interest by time point and by treatment group; these parameters are specified in Appendix 3.
Separate by-subject listings of laboratory data for hematology, biochemistry and urinalysis will be provided by treatment group, with abnormal values highlighted, and including center, subject identifier, age, sex, race, weight and visit. For each subject, all results for any lab parameter that has at least one abnormal value will be included in the listing. Laboratory reference ranges (Lower Limit of Normal, upper Limit of Normal) will be presented for each laboratory parameter.

4.10.9 Vital Signs, Physical Findings and Other Observations Related to Safety

Summaries will be provided for the following vital sign parameters:

Temperature (in °C), systolic and diastolic blood pressure (mmHg), heart rate (beats/min), respiration rate (breaths/min), weight (kg) and calculated BMI (kg/m^2).

Pregnancy testing must be carried out for women of childbearing potential and will consist of serum testing at Screening and urine testing at baseline, Weeks 4, 8, 12, safety follow-up and potential early withdrawal-visit. Pregnancy test results will be listed only.

Findings related to physical examination will be assigned by investigator to medical history, AEs or SAEs.
Subjects with findings or subjects who missed examinations will be shown in a listing.

The following summaries will be provided:

- A summary of each vital sign parameter by treatment group and time point
- A summary of the change from Baseline in each vital sign parameter by treatment group and time point.
- Boxplots of the change from Baseline in each vital sign parameter by treatment group and time point.
- A summary of the number and percentage for history of tobacco use and family history of cardiovascular risk factors by treatment group.

12-lead ECGs will be obtained at screening, Week 12, and potential early withdrawal-visit.

The following summaries will be provided:

- A summary of each ECG parameter (i.e. heart rate, PR, QRS, QT, QTc) by treatment group and time point
- A summary of the change from baseline in each ECG parameter by treatment group and time point.
- A summary of the number and percentage of subjects with abnormal findings for each ECG parameter by treatment group and time point.
- A summary of the number and percentage of subjects with QTc interval exceeding >450ms/ >480ms/ >500ms by treatment group and time point.
- A summary of the number and percentage of subjects with change from baseline in QTcF interval exceeding >30ms/ >60ms by treatment group and time point.
- Boxplots of the change from baseline for each ECG parameter by treatment group and time point.

By-subject listings of vital sign parameters, and ECG results (individual and average) and any other observations related to safety will be provided.

4.10.10 Safety Monitoring

4.10.10.1 Medical Monitoring

Medical monitoring will happen in form of monthly blinded data review meetings. Further details and the list of generated output can be found in the Medical Data Review Plan and the Medical Monitoring Plan.

4.10.11 Immunogenicity

Immunogenicity samples for determination of anti-drug-antibody (ADA) will be collected. Samples taken after dosing with GSK3196165 that have a value at or above the cut-point will be considered potentially treatment-emergent ADA-positive. Shift table from baseline to every assessment will be produced for the Immunogenicity population to assess the number of subjects going from:

1) negative → negative,
2) negative → positive,
3) positive → negative,
4) positive → positive.

Serum analysis will be performed under the management of Immunogenicity and Clinical Immunology (ICI), GlaxoSmithKline. Serum will be tested for the presence of anti-GSK3196165 antibodies using the currently approved analytical methodology incorporating screening, confirmation and titration steps.
Anti-GSK3196165 Binding Antibody Detection (positive/negative) will be listed together with titre value (mL) and Rheumatoid Factor (positive/negative) for each subject with at least one positive result of Anti-GSK3196165 Binding Antibody Detection.

A Listing of DAS28(CRP) values will be produced for all subjects who developed ADA in the study, including RF and ACPA by visit.

4.11 Patient Reported Outcome (PRO) Measures

Change over time and change from baseline in RA Symptom and Impact Diary measures will be summarized for subjects that are in the ITT population. Responses for each of the 16 questions from the RA Symptom and Impact Diary will be summarized by visit.

4.12 Other Analyses

4.12.1 Pharmacokinetic Analysis

Pharmacokinetic concentrations will be listed and summarised by visit.

Pharmacokinetic data may be analysed using population PK approach. Sparse PK samples collected in this study may be pooled with PK NONMEM dataset including all the PK data collected in clinical studies reported so far with GSK3196165. The main objective of this analysis is to derive post-hoc estimates of individual PK parameters characterising the PK time profile of GSK3196165. Further details on the population PK analysis will be provided in a separate technical analysis plan.

4.12.2 Pharmacokinetic/Pharmacodynamic (PK/PD) Exploratory analysis

A graphical PK/PD exploration may be conducted to visualize the nature of the concentration-response correlation. GSK3196165 concentration is characterized by the individual post-hoc estimates from the PK model, and the response is characterized by the individual values of the PD markers (any PD marker identified to be correlated with efficacy) or efficacy (e.g. DAS28(CRP))
Example of PK/PD graphs:

- Linear scatterplot of change from baseline in DAS28(CRP) measures at Week 12 versus Cave estimated from the PK model
- Boxplot of Cave stratified by ACR20 at Week 12
- Boxplot of Cave stratified by ACR50 at Week 12
- Boxplot of Cave stratified by ACR70 at Week 12

4.12.3 Genetic Research

Analyses of genetics will be defined in a separate analysis plan, if deemed necessary.

4.13 Determination of Sample Size

For the MRI synovitis data, based on a review of results in the literature, the standard deviation (SD) for both treatment groups is assumed to have a value of 2.5. Based on this value, assuming prior information cannot be used, with 30 subjects on GSK3196165 180 mg and 10 subjects on placebo, aiming for 50% early and 50% established RA subjects, it is estimated that the lower and upper bounds of the 95% CI for the difference would be within 1.8 points of the difference. This is considered the worst-case precision and further estimates assuming different weights for the prior distribution on placebo will be assessed to estimate any improvement in the precision of the CI.

4.14 Changes in the Conduct of the Study or Planned Analysis

Due to only 9/39 (23%) randomized subjects with early RA (≤2 years disease duration), no subgroup analyses will be performed by disease duration. This may be looked at post-hoc.

The following endpoints will not be summarized or analyzed but will appear in analysis datasets and listing. Post-hoc analyses may be performed.

- ESR value
- DAS28(ESR)
- DAS28(ESR) remission
- DAS28(ESR) LDA
- Index-based remission
- Boolean-based remission
- CDAI remission

5 REFERENCES

[1] Collett D. Modeling Survival Data in Medical Research. 2nd edition. London: Chapman & Hall; 2003.

[2] O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. Biometrics 1979;35;549-56.

[3] Steidl S, Ratsch O, Brocks B, *et al*. In vitro affinity maturation of human GM-CSF antibodies by targeted CDR-diversification. Mol Immunol 2008;46:135-44.

[4] Conaghan P.G, Emery P, Østergaard M, *et al*.. Assessment by MRI of inflammation and damage in rheumatoid arthritis patients with methotrexate inadequate response receiving golimumab: results of the GO-FORWARD trial. Ann Rheum Dis 2011;70:1968-74.

[5] Østergaard M, Emery P, Conaghan PG, *et al*. Significant improvement in synovitis, osteitis, and bone erosion following golimumab and methotrexate combination therapy as compared with methotrexate alone: a magnetic resonance imaging study of 318 methotrexate-naive rheumatoid arthritis patients. Arthritis Rheum 2011;63:3712-22.

[6] Peterfy C, Emery P, Tak PP, *et al*. MRI assessment of suppression of structural damage in patients with rheumatoid arthritis receiving rituximab: results from the randomised, placebo-controlled, double-blind RA-SCORE study. Ann Rheum Dis 2016;75:170-7.

[7] Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population." Am J Respir Crit Care Med 1999;159:179–87.

6 Appendix

6.1 Appendix A: Biomarker Details

Mapping of Biomarker terms to CDISC Controlled Terms will be provided separately.

Target Engagement	Free, soluble GM-CSF GM-CSF:drug complex
Predictive Biomarkers	MRP8/14 complex CCL17 (TARC) 14-3-3□ SAA YKL-40 IL-6 MMP-3 CXCL13 (BLC) CCL22
Complement Biomarkers	C5a TCC C4a C3 C4 sCD163

Cartilage Biomarkers	ARGS Neoepitope C1M C2M C3M CRPM VICM
Mechanistic Biomarkers	TNFα IL-1β IL-8 IL-10 IL-15 IL-17A IL-17F
Flow Cytometry: 6 Colour TBNK Panel	%CD3+ ABSOLUTE CD3+ %CD3+CD8+ ABSOLUTE CD3+CD8+ %CD3+CD4+ ABSOLUTE CD3+CD4+ %CD16+CD56+ ABSOLUTE CD16+ CD56+ %CD19+ ABS CD19+ CELLS ABSOLUTE LYMPH HELPER/SUPPRESSOR
Flow Cytometry: CD16+ Monocyte Panel	CD1br+CD16- / % of WBC CD14br+CD16- /% of Monocytes CD14br+CD16- /Abs in cells/ml CD14br+CD16- / Events CD14br+CD16-HLA-DR+CD200R1+ / % of Classical Monocytes CD14br+CD16-HLA-DR+CD200R1+/ Events CD14lo+CD16br+ / % WBC CD14lo+CD16br+/ % monocytes

	CD14lo+CD16br+ / Abs in cells per ml CD14lo+CD16br+ / Events CD14lo+CD16br+ HLA-DR+ CD200R1+ / % non classical monocytes CD14lo+CD16br+ HLA-DR+ CD200R1+ / Events CD14br+CD16+ / % WBC CD14br+CD16+ / % monocytes CD14br+CD16+ / Abs in cells per ml CD14br+CD16+ / Events CD14br+CD16+ HLA-DR+CD200R1+ / % Int monocytes CD14br+CD16+ HLA-DR+CD200R1+ / Events CD14-HLA-DR+CD11c-CD123br+ / % monocytes CD14-HLA-DR+CD11c-CD123br+ / Abs in cells per ml CD14-HLA-DR+CD11c-CD123br+ / Events CD14-HLA-DR+CD11c- CD123br+ CD200R1+ / % of pDC CD14-HLA-DR+CD11c- CD123br+ CD200R1+ / Events CD14-HLA-DR+CD11c br+ CD123+ / % monocytes CD14-HLA-DR+CD11c br+ CD123+ / Abs in cells per ml CD14-HLA-DR+CD11c br+ CD123+ / Events CD14-HLA-DR+CD11c br+ CD123+ CD200R1+/ % of mDC1 CD14-HLA-DR+CD11c br+ CD123+ CD200R1+/ Events CD14-HLA-DR+CD11c br+ CD123- / % monocytes CD14-HLA-DR+CD11c br+ CD123- / Abs in cells per ml CD14-HLA-DR+CD11c br+ CD123- / Events CD14-HLA-DR+CD11c br+ CD123- CD200R1+ / % of mDC2 CD14-HLA-DR+CD11c br+ CD123- CD200R1+ / Events CD14-CD16+CD66b+ / % WBC CD14-CD16+CD66b+ / per CMM CD14-CD16+CD66b+ / Events CD14-CD16+CD66b+CD200R1+ / % neutrophils CD14-CD16+CD66b+CD200R1+ / Events
Flow Cytometry: T Reg Cell Foxp3	%CD3+ %CD3+CD4+ %CD3+4+25+127- %CD3+4+foxp3+25+127-

	%CD3+CD8+ CD3+A CD3+4+A CD3+4+25+127-A CD3+4+foxp3+25+127-A CD3+8+A
T Helper Cell Panel	CD45+CD3+CD8-CD4+ / % total helper T cells CD45+CD3+CD8-CD4+ / Events total helper T cells CD45+CD3+CD8-CD4+CCR6+CXCR3- (Th17) / % of total helper T cells CD45+CD3+CD8-CD4+CCR6+CXCR3- (Th17) / Events total helper T cells CD45+CD3+CD8-CD4+CCR6+CXCR3-CD38+ HLA-DR+ / % of Th17 cells CD45+CD3+CD8-CD4+CCR6+CXCR3-CD38+ HLA-DR+ / Events of Th17 cells CD45+CD3+CD8-CD4+CCR6-CXCR3- (Th2) / % of total helper T cells CD45+CD3+CD8-CD4+CCR6-CXCR3- (Th2) / Events total helper T cells CD45+CD3+CD8-CD4+CCR6-CXCR3-CD38+ HLA-DR+ / % of Th2 cells CD45+CD3+CD8-CD4+CCR6-CXCR3-CD38+ HLA-DR+ / Events of Th2 cells CD45+CD3+CD8-CD4+CCR6-CXCR3+ (Th1) / % of total helper T cells CD45+CD3+CD8-CD4+CCR6-CXCR3+ (Th1) / Events total helper T cells CD45+CD3+CD8-CD4+CCR6-CXCR3+CD38+ HLA-DR+ / % of Th1 cells CD45+CD3+CD8-CD4+CCR6-CXCR3+CD38+ HLA-DR+ / Events of Th1 cells CD45+CD3+CD8-CD4+CCR6+CXCR3+ (Th proinflam) / % of total helper T cells CD45+CD3+CD8-CD4+CCR6+CXCR3+ (Th proinflam) /Events total helper T cells CD45+CD3+CD8-CD4+CCR6+CXCR3+CD38+ HLA-DR+ / % of Th proinflam cells CD45+CD3+CD8-CD4+CCR6+CXCR3+CD38+ HLA-DR+ / Events of Th proinflam cells
Safety Biomarkers	KL-6 Antigen Surfactant Protein D Anti-GM-CSF auto-antibodies 3-beta-cholestenoic acid

6.2 Appendix B: Laboratory Parameters of Interest

Lab parameters of interest	Grade			
	1	2	3	4
HEMOGLOBIN decrease	Hemoglobin (Hgb) <LLN - 10.0 g/dL; <LLN - 6.2 mmol/L; <LLN - 100 g/L	Hgb <10.0 - 8.0 g/dL; <6.2 - 4.9 mmol/L; <100 - 80g/L	Hgb <8.0 g/dL; <4.9 mmol/L; <80 g/L; transfusion indicated	Life-threatening consequences; urgent intervention indicated
WHITE CELL COUNT decrease	<LLN - 3000/mm3; <LLN - 3.0 x 10e9 /L	<3000 - 2000/mm3; <3.0 - 2.0 x 10e9 /L	<2000 - 1000/mm3; <2.0 - 1.0 x 10e9 /L	<1000/mm3; <1.0 x 10e9 /L
TOTAL NEUTROPHILS ABSOLUTE COUNT	<LLN - 1500/mm3; <LLN - 1.5 x 10e9 /L	<1500 - 1000/mm3; <1.5 - 1.0 x 10e9 /L	<1000 - 500/mm3; <1.0 - 0.5 x 10e9 /L	<500/mm3; <0.5 x 10e9 /L
LYMPHOCYTES ABSOLUTE COUNT decreased	<LLN - 800/mm3; <LLN - 0.8 x 10e9 /L	<800 - 500/mm3; <0.8 - 0.5 x 10e9 /L	<500 - 200/mm3; <0.5 - 0.2 x 10e9 /L	<200/mm3; <0.2 x 10e9 /L
LYMPHOCYTES ABSOLUTE increased		>4000/mm3 - 20,000/mm3 ; >4-2 x 10e9 /L	>20,000/mm3; >20 x 10e9 /L	
PLATELET COUNT	<LLN - 75,000/mm3; <LLN - 75.0 x 10e9 /L	<75,000 - 50,000/mm3; <75.0 - 50.0 x 10e9 /L	<50,000 - 25,000/mm3; <50.0 - 25.0 x 10e9 /L	<25,000/mm3; <25.0 x 10e9 /L
CREATININE	>1 - 1.5 x baseline; >ULN - 1.5 x ULN	>1.5 - 3.0 x baseline; >1.5 - 3.0 x ULN	>3.0 baseline; >3.0 - 6.0 x ULN x ULN	>6.0 x ULN
SODIUM decrease	<LLN - 130 mmol/L		<130 - 120 mmol/L	<120 mmol/L;
SODIUM increase	>ULN - 150 mmol/L	>150 - 155 mmol/L	>155 - 160 mmol/L	>160 mmol/L
POTASSIUM decrease	<LLN - 3.0 mmol/L	<LLN - 3.0 mmol/L	<3.0 - 2.5 mmol/L	<2.5 mmol/L
POTASSIUM increase	>ULN - 5.5 mmol/L	>5.5 - 6.0 mmol/L	>6.0 - 7.0 mmol/L	>7.0 mmol/L
CALCIUM increase	>ULN - 11.5 mg/dL; >ULN - 2.9 mmol/L	>11.5 - 12.5 mg/dL; >2.9 - 3.1 mmol/L	>12.5 - 13.5 mg/dL; >3.1 - 3.4 mmol/L;	>13.5 mg/dL; >3.4 mmol/L

CALCIUM decrease	<LLN - 1.0 mmol/L	<1.0 - 0.9 mmol/L	<0.9 - 0.8 mmol/L	<0.8 mmol/L
PHOSPHORUS INORGANIC	<LLN - 2.5 mg/dL; <LLN - 0.8 mmol/L	<2.5 - 2.0 mg/dL; <0.8 - 0.6 mmol/L	<2.0 - 1.0 mg/dL; <0.6 - 0.3 mmol/L	<1.0 mg/dL; <0.3 mmol/L
ASAT (SGOT)	>ULN - 3.0 x ULN	>3.0 - 5.0 x ULN	>5.0 - 20.0 x ULN	>20.0 x ULN
ALAT (SGPT)	>ULN - 3.0 x ULN	>3.0 - 5.0 x ULN	>5.0 - 20.0 x ULN	>20.0 x ULN
GGT	>ULN - 2.5 x ULN	>2.5 - 5.0 x ULN	>5.0 - 20.0 x ULN	>20.0 x ULN
ALKALINE PHOSPHATASE	>ULN - 2.5 x ULN	>2.5 - 5.0 x ULN	>5.0 - 20.0 x ULN	>20.0 x ULN
BILIRUBIN, TOTAL	>ULN - 1.5 x ULN	>1.5 - 3.0 x ULN	>3.0 - 10.0 x ULN	>10.0 x ULN
CPK, TOTAL	>ULN - 2.5 x ULN	>2.5 x ULN - 5 x ULN	>5 x ULN - 10 x ULN	>10 x ULN
ALBUMIN	<LLN - 3 g/dL; <LLN - 30 g/L	<3 - 2 g/dL; <30 - 20 g/L	<2 g/dL; <20 g/L	
GLUCOSE	>ULN -160 mg/dL; >ULN - 8.9 mmol/L	>160 -250 mg/dL; >8.9 - 13.9 mmol/L	>250 - 500 mg/dL; >13.9 - 27.8	>500 mg/dL; >27.8 mmol/L
EST.CREATININE CLEARANCE	<LLN - 60 ml/min/1.73 m2	59 - 30 ml/min/1.73 m2	29 - 15 ml/min/1.73 m2	<15 ml/min/1.73 m2

6.3 Appendix C: Bayesian Repeated Measures Model

6.3.1 Proc Mixed

The modelling will be described using PROC MIXED nomenclature, since it may be more familiar to SAS programmers than the equivalent PROC MCMC statements.

The PROC MIXED estimates and CIs would be labelled using their Bayesian equivalents (e.g. adjusted median and credible intervals).

Subject, treatment and visit will be fitted as class variables. Baseline will be fitted as a continuous effect. All main effects and interaction terms for Treatment and visit will be fitted as fixed effects. A Type=UN repeated measures variance covariance structure will be attempted (with repeated measures on visit and subject=subject).

The following is an example extract of the SAS PROC MIXED code (model terms may be altered if the model experiences fitting/convergence problems – SAS code to extract relevant model results / outputs for model checking is not included below but should be added):

```
proc mixed data=<data>;
class <subject> <treatment> <visit>;
model <response> = <treatment>|<visit> <baseline> * <visit> / ddfm=kr;
repeated <visit>/ subject=<subject> type=UN;
lsmeans <treatment>*<visit> / diff cl;
run;
```

In PROC MIXED nomenclature the adjusted least square mean estimates and their 95% CIs should be extracted. The difference to placebo (within each of the visits) for treatment arm should also be extracted, along with their respective 95% CIs.

The adjusted point estimates for each visit and treatment combination and the adjusted point estimates for the difference to placebo per visit (and associated 95% credible intervals) will be summarized and plotted graphically. For the figure of adjusted medians by visit the individual subject responses should be displayed alongside the corresponding point estimate and credible interval.

If PROC MCMC is used then the suggested model fitting tips and convergence checks described in Section 6.3.2 should be applied.

Transformations of the response variable and/or alternative analysis techniques will be explored on an as needed basis if model assumptions appear invalid (these techniques would be documented in the CSR).

6.3.2 General Considerations for Bayesian Analyses (PROC MCMC)

The following points are for guidance and illustration and do not guarantee a successful model convergence. They cannot cover all eventualities and do not remove the requirement to do what is best for the specific set of observed data being modelled.

Priors

Unless otherwise specified the following would be the default approach to selecting prior distributions:

- Non-informative priors of the form Normal(0, Var=1E6) would be assigned to each fixed effect in the proposed statistical model.
- Non-informative Inverse-gamma priors of the form IGamma(0.001,0.001) would be assigned to stand-alone variance parameters that are not expected to take values near zero (e.g. for the residual variance rather than a random effect variance component)
- For stand-alone variance parameters that may take values close to zero a noninformative prior of the form Uniform(0, XXX) may be assigned for the SD
- For repeated measures models non-informative Inverse-Wishart priors will be assigned for the Variance Covariance matrix (VCV). They would use degrees of freedom equal to the dimension of the VCV matrix and an Identity matrix (of the same dimension).
    - If there are issues with those variance parameters then the Identitity matrix may be replaced with a diagonal matrix that uses best guesses for the residual variance at each repeated measure time point (or the residual estimates from fitting simple models).
    - It is good practice to ensure that each prior distribution is visualised to ensure it appears sensible and allows clinically plausible response values, whilst not allowing impossible values to be drawn with high probability and that if it is intended to be non-informative it is doing so over the region of the likelihood function. Note: There is no requirement to formally report these visualisation outputs.

Power Prior

If $\pi_0(\theta)$ is a non-informative prior defined suitably as above, and $L(\theta|D_0)$ is the distribution of the historical data, then the power prior is given by the following formula:

$$\pi(\theta|D_0) \propto L(\theta|D_0)^{\alpha_0}\pi_0(\theta)$$

Where α_0 is the weight of the power prior.

Initial Values

Unless otherwise specified initial parameter values of zero would be used for the fixed effect parameters and for remaining model parameters initial values may be drawn at random from their respective prior distribution. If model converge is problematic then alternative estimates, more suited to the particular dataset being modelled, may be used

(for example, these could be based on visual inspection of the raw data and/or parameter estimates from fitting simpler models).

Checking convergence and other diagnostics

The key model diagnostic output is the MCSE/SD ratio for each parameter:

- Adequate values for the number of MCMC samples / thinning / number of burn in samples should be chosen to ensure that the MCSE/SD for the key parameters is below 0.01 (e.g. key parameters those associated with treatment, or as pre-specified comparisons of interest) in each final model.
- For other model parameters, try to get the MCSE/SD values as close to 0.01 as possible, but if there is significant autocorrelation then values below 0.05 would be considered acceptable.
- In addition, the number of tuning units and maximum number of tuning iterations may be increased to find a better multivariate normal approximation to the parameters, which in turn may reduce the MCSE/SD values
- Where possible the code should be written to allow the SAS compiler to identify and use conjugate sampling, since this can greatly reduce the corresponding MCSE/SD values
- Models selected with MCSE/SD values above 0.01 (for key parameters) or 0.05(for other parameters) would need a brief remark/justification added to the CSR to clarify why it was not possible to reach the targets and why it is believed the subsequent model still has utility.

Use of the default SAS PROC MCMC diagnostic plots should be made and where possible:

- Autocorrelation should decline rapidly and show no oscillation patterns
- Worm plots should show the chain appears to be stationary and mixing, i.e.
  - Constant mean, constant variance
  - Moving around the parameter space freely (not getting "stuck" at similar values for a large number of iterations before moving on again)
  - Moving rapidly between extremes
  - The posterior density should look reasonable for each parameter (e.g. for posterior parameters expected to follow a normal distribution the density plot should not appear bi-modal, but for parameters acting as binary flags then bi-modal is acceptable)
  - Correlation structures between relevant posterior parameters (and/or parameters in each PARMS block) should be explored using graphical methods. This can provide information about what potential issues may be and also what corrective action(s) may be worthwhile attempting.

Possible corrective actions

Possible corrective actions include but are not limited to:

- Moving parameters (or combinations of parameters) onto separate PARMS statements (to form blocks that are updated independently in the MCMC algorithms)
- Increasing the number of MCMC draws from the posterior distributions

- Increasing the length of the burn in period
- Increasing the thinning parameter (to reduce the autocorrelation)
- Centring covariates (to reduce correlations between the posterior parameters)
- Re-parameterising the model (e.g. using log-normal prior distributions to stop values being sampled that are below zero)
- Re-scaling the parameters (e.g. if one parameter takes values orders of magnitudes different to the other model parameters then dividing it by a suitable constant but back transforming the rescaled parameter prior to its use in any subsequent manipulations)
- Visualising the likelihood function of the dataset to determine if there are more appropriate starting estimates (or profiled versions of the likelihood for subsets of model parameters where difficulties are being encountered if it is a high dimensionality problem)

Utilising the posterior distributions

Equi-tail credible intervals should be used, since they are invariant under transformation of the parameter space (unlike Highest Posterior Density), and it is expected that the models to biomarker data will use a log-transformation, but the quantities of interest will be back transformed (hence transformations of the parameter space will occur in this study).

6.4 Appendix D: Opportunistic Infections: MedDRA Preferred Terms

Opportunistic Infections: MedDRA Preferred Terms	
AIDS retinopathy	Herpes pharyngitis
Acid fast bacilli infection	Herpes sepsis
Acinetobacter bacteraemia	Herpes simplex
Acinetobacter infection	Herpes simplex cervicitis
Acquired immunodeficiency syndrome	Herpes simplex hepatitis
Actinomycosis	Herpes simplex meningoencephalitis
Actinomycotic abdominal infection	Herpes simplex otitis externa
Actinomycotic pulmonary infection	Herpes simplex test positive
Acute HIV infection	Herpes simplex virus conjunctivitis neonatal
Acute hepatitis B	Herpes simplex visceral
Acute hepatitis C	Herpes virus infection
Adrenal gland tuberculosis	Herpes zoster
Arthritis fungal	Herpes zoster cutaneous disseminated
Arthritis salmonella	Herpes zoster disseminated
Aspergilloma	Herpes zoster infection neurological
Aspergillosis oral	Herpes zoster meningitis
Aspergillus infection	Herpes zoster meningoencephalitis
Asymptomatic HIV infection	Herpes zoster meningomyelitis
Asymptomatic viral hepatitis	Herpes zoster necrotising retinopathy
Atypical mycobacterial infection	Herpes zoster oticus
Atypical mycobacterial lower respiratory tract infection	Herpes zoster pharyngitis
Atypical mycobacterial lymphadenitis	Histoplasmosis
Atypical mycobacterial pneumonia	Histoplasmosis cutaneous
Atypical mycobacterium pericarditis	Histoplasmosis disseminated
Bacterial parotitis	Human T-cell lymphocytic virus type II infection
BK virus infection	Human T-cell lymphotropic virus infection
Biliary tract infection cryptosporidial	Human T-cell lymphotropic virus type I infection
Biliary tract infection fungal	Human polyomavirus infection
Blastomycosis	Immune reconstitution inflammatory syndrome associated tuberculosis
Bone tuberculosis	Intestinal tuberculosis
Brachyspira infection	JC virus granule cell neuronopathy
Brain empyema	JC virus infection
Bronchitis fungal	Joint tuberculosis

Opportunistic Infections: MedDRA Preferred Terms	
Bronchopulmonary aspergillosis	Kaposi's sarcoma
Bronchopulmonary aspergillosis allergic	Kaposi's sarcoma AIDS related
Brucella sepsis	Kaposi's varicelliform eruption
Brucellosis	Leptotrichia infection
Candida endophthalmitis	Listeria encephalitis
Candida infection	Listeria sepsis
Candida osteomyelitis	Listeriosis
Candida pneumonia	Lower respiratory tract herpes infection
Candida retinitis	Lower respiratory tract infection fungal
Candida sepsis	Lymph node tuberculosis
Cerebral aspergillosis	Lymphadenitis fungal
Cerebral fungal infection	Lymphoma AIDS related
Cerebral toxoplasmosis	Male genital tract tuberculosis
Choroid tubercles	Meningitis aspergillus
Chronic hepatitis	Meningitis candida
Chronic hepatitis C	Meningitis coccidioides
Coccidioides encephalitis	Meningitis cryptococcal
Coccidioidomycosis	Meningitis fungal
Colitis herpes	Meningitis herpes
Congenital HIV infection	Meningitis histoplasma
Congenital cytomegalovirus infection	Meningitis listeria
Congenital hepatitis B infection	Meningitis salmonella
Congenital herpes simplex infection	Meningitis toxoplasmal
Congenital toxoplasmosis	Meningitis tuberculous
Congenital tuberculosis	Meningoencephalitis herpes simplex neonatal
Congenital varicella infection	Meningoencephalitis herpetic
Conjunctivitis tuberculous	Minor cognitive motor disorder
Corynebacterium infection	Mucocutaneous candidiasis
Corynebacterium sepsis	Mycobacterial infection
Cryptococcal cutaneous infection	Mycobacterium abscessus infection
Cryptococcal fungaemia	Mycobacterium avium complex immune restoration disease
Cryptococcosis	Mycobacterium avium complex infection
Cryptosporidiosis infection	Mycobacterium chelonae infection
Cutaneous coccidioidomycosis	Mycobacterium fortuitum infection
Cutaneous tuberculosis	Mycobacterium kansasii infection
Cytomegalovirus chorioretinitis	Mycobacterium marinum infection
Cytomegalovirus colitis	Mycobacterium ulcerans infection
Cytomegalovirus duodenitis	Mycotic endophthalmitis
Cytomegalovirus enteritis	Mycotoxicosis

Opportunistic Infections: MedDRA Preferred Terms	
Cytomegalovirus enterocolitis	Myocarditis mycotic
Cytomegalovirus gastritis	Myocarditis toxoplasmal
Cytomegalovirus gastroenteritis	Nasal herpes
Cytomegalovirus gastrointestinal infection	Necrotising fasciitis fungal
Cytomegalovirus gastrointestinal ulcer	Necrotising herpetic retinopathy
Cytomegalovirus infection	Neonatal candida infection
Cytomegalovirus mononucleosis	Neonatal mucocutaneous herpes simplex
Cytomegalovirus mucocutaneous ulcer	Neurocryptococcosis
Cytomegalovirus myelomeningoradiculitis	Neutropenic sepsis
Cytomegalovirus myocarditis	Nocardia sepsis
Cytomegalovirus oesophagitis	Nocardiosis
Cytomegalovirus pancreatitis	Oesophageal candidiasis
Cytomegalovirus pericarditis	Oesophageal tuberculosis
Cytomegalovirus syndrome	Ophthalmic herpes simplex
Cytomegalovirus urinary tract infection	Ophthalmic herpes zoster
Cytomegalovirus viraemia	Oro-pharyngeal aspergillosis
Disseminated cryptococcosis	Oropharyngeal candidiasis
Disseminated cytomegaloviral infection	Osteomyelitis blastomyces
Disseminated tuberculosis	Osteomyelitis fungal
Ear tuberculosis	Osteomyelitis salmonella
Eczema herpeticum	Overgrowth fungal
Encephalitis cytomegalovirus	Pancreatitis fungal
Encephalitis fungal	Paratyphoid fever
Encephalitis post immunisation	Pericarditis fungal
Encephalitis post varicella	Pericarditis histoplasma
End stage AIDS	Pericarditis tuberculous
Endocarditis candida	Perinatal HIV infection
Endocarditis histoplasma	Peritoneal candidiasis
Enterocolitis AIDS	Peritoneal tuberculosis
Enterocolitis fungal	Persistent generalised lymphadenopathy
Epididymitis blastomyces	Pneumocystis jirovecii infection
Epididymitis tuberculous	Pneumocystis jirovecii pneumonia
Epstein-Barr virus associated lymphoma	Pneumonia blastomyces
Erythema induratum	Pneumonia cytomegaloviral
Erythrasma	Pneumonia fungal
Exanthema subitum	Pneumonia herpes viral
Extrapulmonary tuberculosis	Pneumonia salmonella
Eye infection toxoplasmal	Pneumonia toxoplasmal
Female genital tract tuberculosis	Polyomavirus-associated nephropathy
Fungaemia	Presumed ocular histoplasmosis syndrome

Opportunistic Infections: MedDRA Preferred Terms	
Fungal abscess central nervous system	Proctitis herpes
Fungal endocarditis	Proctitis monilial
Fungal infection	Progressive multifocal leukoencephalopathy
Fungal oesophagitis	Prostatitis tuberculous
Fungal peritonitis	Pulmonary mycosis
Fungal sepsis	Pulmonary trichosporonosis
Fungal tracheitis	Pulmonary tuberculoma
Funguria	Pulmonary tuberculosis
Gastritis fungal	Pyelonephritis fungal
Gastritis herpes	Renal tuberculosis
Gastroenteritis cryptococcal	Respiratory moniliasis
Gastroenteritis cryptosporidial	Respiratory tract infection fungal
Gastrointestinal candidiasis	Retinitis histoplasma
Gastrointestinal fungal infection	Retroviral infection
Genital blister	Retroviral rebound syndrome
Genital herpes	Salmonella bacteraemia
Genital herpes zoster	Salmonella sepsis
Haemorrhagic pneumonia	Salmonellosis
Hepatosplenic abscess	Salpingitis tuberculous
HIV associated nephropathy	Silicotuberculosis
HIV cardiomyopathy	Sinusitis aspergillus
HIV enteropathy	Skin candida
HIV infection	Spleen tuberculosis
HIV infection CDC Group I	Splenic candidiasis
HIV infection CDC Group II	Splenic infection fungal
HIV infection CDC Group III	Stoma site candida
HIV infection CDC Group IV subgroup A	Stoma site infection
HIV infection CDC Group IV subgroup B	Superinfection fungal
HIV infection CDC Group IV subgroup C1	Superinfection mycobacterial
HIV infection CDC Group IV subgroup C2	Systemic candida
HIV infection CDC Group IV subgroup D	Systemic mycosis
HIV infection CDC Group IV subgroup E	T-cell lymphoma
HIV infection CDC category A	T-cell type acute leukaemia
HIV infection CDC category B	Thyroid tuberculosis
HIV infection CDC category C	Tongue fungal infection
HIV infection CDC group IV	Toxoplasmosis
HIV infection WHO clinical stage I	Tropical spastic paresis
HIV infection WHO clinical stage II	Tuberculoma of central nervous system
HIV infection WHO clinical stage III	Tuberculosis

Opportunistic Infections: MedDRA Preferred Terms	
HIV infection WHO clinical stage IV	Tuberculosis bladder
HIV peripheral neuropathy	Tuberculosis gastrointestinal
HIV wasting syndrome	Tuberculosis liver
Hepatic candidiasis	Tuberculosis of central nervous system
Hepatic infection fungal	Tuberculosis of eye
Hepatitis B	Tuberculosis of genitourinary system
Hepatitis C	Tuberculosis of intrathoracic lymph nodes
Hepatitis chronic persistent	Tuberculosis of peripheral lymph nodes
Hepatitis fulminant	Tuberculosis ureter
Hepatitis toxoplasmal	Tuberculous abscess central nervous system
Hepatitis viral	Tuberculous endometritis
Hepatitis virus-associated nephropathy	Tuberculous laryngitis
Hepatosplenic candidiasis	Tuberculous pleurisy
Herpes dermatitis	Tuberculous tenosynovitis
Herpes oesophagitis	Typhoid fever
Herpes ophthalmic	Varicella keratitis

6.5 Appendix E: Scoring of RAMRIS Features

Synovitis Scoring

Score	Description
0	Normal, no synovitis
1	1-33% volume enhancement
2	34-66% volume enhancement
3	67-100% volume enhancement
N	Not Visible
S	Surgically Modified

Bone Erosion Scoring

Score	Description
0	No erosion
1	1-10% of bone eroded
2	11-20% of bone eroded
3	21-30% of bone eroded
4	31-40% of bone eroded
5	41-50% of bone eroded
6	51-60% of bone eroded
7	61-70% of bone eroded
8	71-80% of bone eroded
9	81-90% of bone eroded
*** (10)**	91-100% of bone eroded
N	Not Visible
A	Bones with ankylosis of >50% of their surfaces; Not Assessable
S	Surgically Modified

Bone Edema/Osteitis Scoring

Score	Description
0	No bone edema/osteitis
1	1-33% involvement of original articular bone
2	34-66% involvement of original articular bone
3	67-100% involvement of original articular bone
N	Not Visible
S	Surgically Modified

Cartilage Loss Scoring

Score	Description
0	No cartilage loss or Joint Space Narrowing (JSN)
0.5	Equivocal cartilage loss or JSN
1	Minimal (< 10%) but definitive cartilage loss or JSN
1.5	Mild (10% to 25%) cartilage loss or JSN
2	Moderate cartilage loss or JSN (26% to 75%, including unilaterally denuded areas but no bilaterally denuded areas or bone-on-bone contact)

2.5	Moderate-severe cartilage loss or JSN (> 75%, including focal denuding or focal bone-on-bone contact)
3	Complete cartilage denuding or diffuse bone-on-bone contact
3.5	Partial ankylosis
4	Complete ankylosis
N	Not Visible
S	Surgically Modified

6.6 Appendix F: RAP Updates

Version 2.0

Sponsor Signature Page
Update as per sponsor request.

Sections 3.3.12, 3.3.13, 3.3.14
Details of derivation of MRI endpoints have been corrected.
Imputation Rules have been updated.

Section 4.4.4
Section about rules for visit windows and derivation of composite efficacy scores has been added.

Section 4.9.2
Statistical methods to be used for analysis of exploratory biomarkers have been described in more detail.

Section 4.9.3.1
Statistical methods to be used for analysis of MRI endpoints have been described in more detail.

Section 4.9.3.3
Statistical methods to be used for analysis of clinical efficacy endpoints have been described in more detail. Table 4 in Section 4.9.4 has been updated accordingly.

Section 4.10.5
Details for reporting of spirometry results have been added.

Section 6.1 Appendix F
List of biomarkers has been updated. New category for safety biomarkers has been added. Section 4.9.2 and Table 5 in Section 4.9.4 have been updated accordingly.

Document ID: 393172927 - Name at Publication: 227221 RAP 20171116 2.0
This document contains a total of 71 pages including any signature page/s.

PAREXEL International Electronic Signature Page

This page is the manifestation of the electronic signature(s) used in compliance with PAREXEL International's electronic signature policies and procedures and in compliance with applicable regulations.

UserName: PPD
Title: Biostatistician II, GRO
Date: Wednesday, 29 November 2017, 06:26 AM GMT Standard Time
Meaning: Document contents authored.
==========

UserName: PPD
Title: Senior Medical Writer, GMS
Date: Wednesday, 29 November 2017, 06:29 AM GMT Standard Time
Meaning: Document contents approved.
==========

UserName: PPD
Title: Principal Statistical Programmer, GRO
Date: Wednesday, 29 November 2017, 06:31 AM GMT Standard Time
Meaning: Document contents approved.
==========

UserName: PPD
Title: Senior Biostatistician, GRO
Date: Wednesday, 29 November 2017, 08:57 AM GMT Standard Time
Meaning: Document contents approved.
==========