Clinical Trials Research Unit
(CTRU)
University of Leeds

Statistical Analysis Plan

SWITCH

Version 1.0

4th November 2015

Trial Statistician:	Colin Everett
Supervising Trial Statistician:	Sarah Brown
Study Scientific Lead	Linda Sharples
Project Delivery Lead	Catherine Fernandez
Senior Trial Co-ordinator:	Claire Davies
Data Manager:	Catherine Reynolds
Chief Investigator:	Maya Buch

Glossary

ABT	Abatacept. One of the study drugs in the SWITCH study.
ACPA	Anti-Citrullinated Peptide Antibody – the assay that detects presence of anti-CCP
ACR	American College of Rheumatology
ADA	Adalimumab. One of the allowable monoclonal antibody treatments under the SWITCH study
Anti-CCP	Anti-Cyclic Citrullinated Peptide.
Anti-TNF	See TNFi
Arthritis Research UK AIA CSG	Arthritis Research United Kingdom AIA Clinical Studies Group. An arthritis-related special interest group.
BSR	British Society for Rheumatology
CC	Complete Case (analysis). A patient with complete data for all fields required in the analysis.
CDAI	Clinical Disease Activity Index
CONSORT	Consolidated Standards of Reporting Trials. Refers to either the patient flow diagram, recommended by such guidance, or the guidance itself. See references (1,2) and Appendix A
CRF	Case Report Form
CRP	C-Reactive Protein. A measure of inflammation.
CTRU	Clinical Trials Research Unit, Leeds Institute of Clinical Trials Research, University of Leeds
CTZ	Certoliuzumab pegol. One of the allowable monoclonal antibody treatments under the SWITCH study
DAS28	Disease Activity Score with 28 joint counts. A composite outcome measure for patients with Rheumatoid Arthritis. (See Section 2.3 for definition)
DCF	Data Clarification Form
DMA	Data Management Assistant
DMARD	Disease-modifying Anti-Rheumatic Drug
DMEC	Data Monitoring and Ethics Committee
EQ5D	EuroQol 5-dimensions questionnaire.

ESR	Erythrocyte Sedimentation Rate. A measure of inflammation.
ETN	Etanercept. One of the possible alternative TNFi options in the SWITCH study.
EULAR	European League Against Rheumatism
GCP	Good Clinical Practice
GOL	Golimumab. One of the allowable monoclonal antibody treatments under the SWITCH study
HADS	Hospital Anxiety and Depression Scale. A Quality of Life questionnaire.
HAQ-DI	Health Assessment Questionnaire – Disability Index. A Quality of Life questionnaire.
HTA	Health Technology Assessment. The funding stream for the SWITCH study.
IB	Investigator Brochure
ICC	Intra-class correlation coefficient
IFX	Infliximab. One of the allowable monoclonal antibody treatments under the SWITCH study
ITT	Intention to Treat
IV	Intravenous
LDA	Low Disease Activity
MAB	Monoclonal Antibody. One of the possible alternative TNFi options in the SWITCH study
MAR	Missing At Random. The assumption that if a data item is missing, the "missingness" is not related to its underlying unobserved value once we account for the observed values of other variables in the imputation model.
MTX	Methotrexate. A required concomitant medication for patients in SWITCH.
NHS	National Health Service
NICE	National Institute for Health and Clinical Excellence
NSAID	Non-steroidal anti-inflammatory drug
NYHA	New York Heart Association

QoL	Quality of Life
RA	Rheumatoid Arthritis
RAQoL	Rheumatoid Arthritis Quality of Life. A Quality of Life Questionnaire
REFLEX	Rituximab for Rheumatoid Arthritis Refractory to Anti–Tumor Necrosis Factor Therapy. 2006 Phase III randomised controlled trial assessing the benefit of rituximab vs placebo in patients receiving methotrexate who failed their initial TNFi therapy. See reference (3).
RF	Rheumatoid Factor
RTX	Rituximab – One of the study drugs in the SWITCH study
SAS	Statistical Analysis Software. Cary NC, USA
SDAI	Simplified Disease Activity Index
SJC	Swollen Joint Count. The number of joints out of 28 that are swollen.
SmPC	Summary of Product Characteristics
SUSAR	Suspected Unexpected Serious Adverse Reaction
SWITCH	The present study. For protocol paper, see reference (4)
TB	Tuberculosis
TC	Trial Co-ordinator
TJC	Tender Joint Count. The number of joints out of 28 that are tender.
TMG	Trial Management Group
TNF (alpha)	Tumor Necrosis Factor (alpha). A biomarker indicative of inflammation.
TNFi	Tumor Necrosis Factor (alpha) Inhibitor. An agent that acts to reduce levels of this biomarker.
TSC	Trial Steering Committee
VAS	Visual Analogue Scale. A means of assessing a patient-reported outcome.
WCBP	Woman of Child Bearing Potential

Box 1: Primary Endpoint: DAS28

$$DAS28 = 0.56 \times \sqrt{TJC} + 0.28 \times \sqrt{SJC} + 0.7 \times log_e ESR + 0.014 \times VAS$$

Where:

TJC / SJC = Tender (or Swollen) Joint Counts

ESR = Erythrocyte Sedimentation Rate

VAS = Patient Completed Visual Analogue Scale Score of Global Assessment of Arthritis (mm)

1. Introduction

1.1 Background

Rheumatoid arthritis (RA) is one of the most common autoimmune diseases; a chronic, systemic, inflammatory arthritis, affecting over 600,000 people in the UK (5) and is the largest cause of treatable disability in the Western world (6,7). Patients suffer considerable pain, stiffness and swelling and if not adequately controlled, sustain various degrees of joint destruction, deformity, and significant functional decline. RA can occur at any age, but the peak time of onset is in the fourth and fifth decades of life, a time which coincides with i) highest earning potential for those in work, and ii) may also represent a significant transition phase in roles within the family -including dealing with adolescents moving toward independence at one end of the spectrum and likely increased dependence of elderly parents at the other end (8).

Given its high prevalence in the working population, the impact of RA represents a major individual and societal economic burden (9). The significant direct costs of hospitalisation, joint replacement surgery, drugs and social care are matched with equivalent indirect financial impact, through loss of employment. Expedient implementation of disease-modifying anti-rheumatic drug (DMARD) therapy is the cornerstone of management of RA. Nevertheless, it has become clear that poor response (even if initially effective) remains a feature with most DMARDs over time. In addition, a high incidence of toxicity has been observed with these drugs (10). Such obstacles to therapy combined with data suggesting limited alteration in long-term outcome even in those showing response has argued for more optimal therapy (11).

This unmet clinical need fuelled research into RA which led to significant advances in our understanding of RA by the 1990s; excess pro-inflammatory cytokines, in particular, TNF-alpha was shown to be critical in driving RA pathogenesis (12). Following in vitro and in vivo work, the most compelling evidence for a key role for TNF-alpha stemmed from studies where marked clinical benefit was observed in patients with RA treated with chimeric anti-TNF-alpha monoclonal antibodies (13).

1.1.1 TNF-Inhibitors

Cochrane reviews provide clear evidence that three currently licensed TNFi drugs (etanercept, infliximab and adalimumab) produce better outcomes in RA compared with placebo or treatment with conventional DMARDs (14). All these are in the same class of drug i.e. TNFi but differ in several respects, such as molecule type, target, binding affinity to TNFi, mechanism or action and method of administration.

Despite the extensive benefits of TNF-directed biologic therapies, a significant proportion of RA patients fail to achieve sufficient response (15). Two broad approaches can be employed to manage initial TNFi non-response; switching to an alternative TNFi therapy or use of another mechanism agent. Of the latter, rituximab, a B-cell depleting therapy and abatacept, another agent that targets T-cell co-stimulation are licensed, with rituximab also approved by National Institute of Health and Clinical Excellence (NICE) for the treatment of RA.

Tocilizumab, an interleukin-6 receptor monoclonal antibody, has also been recently licensed and approved by NICE following TNFi failure.

1.1.2 Switching between TNF-Inhibitors

Current NICE guidance does not permit switching to an alternative TNFi as a second-line biologic therapy choice. Several early phase, uncontrolled studies and an initial small randomised study suggested benefit in switching between TNFi agents (16-26). The rationale and argument for switching between different anti-TNF drugs was recently strengthened by a large, randomised industry-led efficacy study comparing Golimumab to Placebo in a Phase II study of 461 patients previously having failed or intolerant to 1 or more TNF-inhibitors (27). A key benefit of the TNFi is their suitability in both seropositive and seronegative disease (to rheumatoid factor (RF) +/- anti-citrullinated peptide antibody (ACPA)). This contrasts with the influence of antibody status and response rates in patients treated with rituximab due to its distinct target and rationale for use (rituximab depletes B-cells that produce antibodies; see below). It is important therefore not to prematurely discount an alternative TNFi drug as an effective therapeutic option, particularly in the context of such resistant and aggressive disease cohorts.

1.1.3 Alternative Biologic Therapies

Recently introduced alternative targeted biological therapies provide another option in the setting of TNFi failure. These include rituximab and abatacept. Industry-led efficacy studies have demonstrated benefits of both these therapies after TNFi failure (3, 28) although only rituximab is NICE-approved (and neither abatacept nor a TNF-antagonist has been compared to rituximab). Certain patients however will not be appropriate for rituximab (and may even lead to unpredictable responses/toxicity (29)) or will fail to respond (up to a third of patients). Furthermore, seronegative antibody status (seen in up to 25-30% of patients in this cohort) is associated with poorer response although this has not been formally tested (3, 30, 31). Abatacept's mechanism, like the TNFi therapies is associated with use in both seropositive and seronegative RA.

A recently published Swiss observational study (32) comprised 116 patients that had failed at least one TNFi agent that were either switched to an alternative TNFi therapy or to one cycle of rituximab with suggestion of rituximab a more favourable treatment option. Aside from including small numbers, this retrospective study had several other design limitations with outcome taken from differing time-points and inclusion of all types of initial TNFi failure; in addition it was neither controlled nor randomised to treatment type. We recently reported an interim observational analysis of patients switched to either an alternative TNFi or rituximab following failure of one/more TNFi therapies; this suggested equivalent clinical responses (33). Similar conclusions were drawn from another Swiss study (34). Notably, meta-analyses have failed to demonstrate superiority of one therapy over another (35), with European recommendations also confirming all as appropriate options (36).

Despite the benefits of recent advances in the management of RA, it is also apparent that no universally effective treatment exists. It remains unclear how best to utilise the alternative biological therapies described above following initial TNFi failure. The present approach is unsatisfactory, with clinicians treating patients in the absence of sufficiently strong data. The current reality, of 2nd-line biologic treatment restricted to a single option (rituximab) seriously impedes effective management. This is particularly pertinent to TNFi failure patients that have seronegative RA (up to 25-30% patients) for whom no NICE-approved options exist despite several more appropriate licensed therapies available and indeed other pathologies or disease characteristics that would argue for an alternative line of management. This poses a significant problem to the NHS and is in conflict with the patient agenda. Despite several treatment options now available, no

good quality head-to-head comparisons investigating the efficacy of sequential biologic treatments have been conducted to date.

In the re-design, 477 patients would be allocated on a 1:1:1 ratio to either rituximab (RTX), abatacept (ABT), or alternative TNFi (SWITCH). Within the alternative TNFi arm, a patient previously failing to respond to a monoclonal antibody will receive etanercept (ETN) and a patient failing to respond to etanercept will receive a monoclonal antibody (MAB) at the discretion of the treating clinician. Possible Monoclonal antibodies will include certolizumab (CTZ), golimumab (GOL), infliximab (IFX) or adalimumab (ADA). Following early trial closure, only 122 patients were randomised.

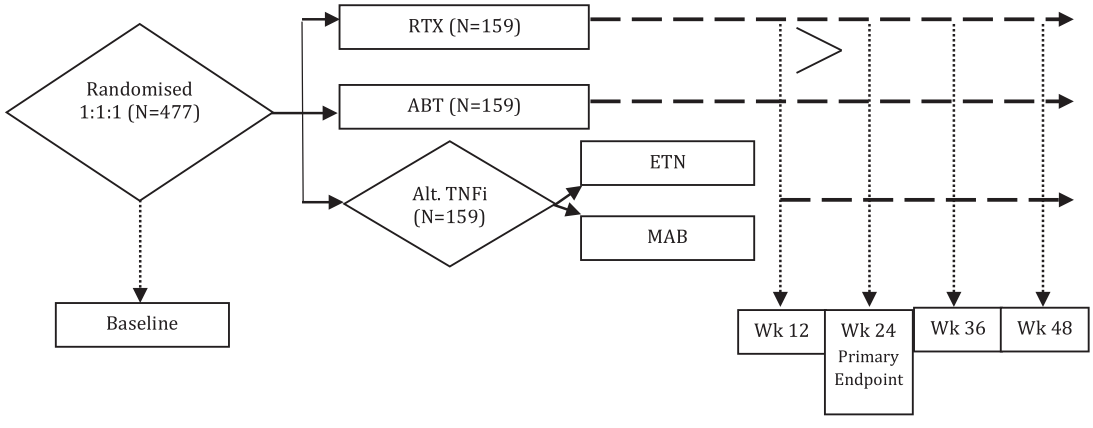


Figure 1: Brief trial design.

1.2 Design

1.2.1 Current Trial Design

SWITCH is a UK multi-centre, Phase IV 3-arm parallel group, randomised controlled trial. A total of 477 patients with Rheumatoid Arthritis, receiving Methotrexate and having failed to respond to initial TNFi therapy will be randomised to receive one of rituximab, abatacept, or alternative TNFi. Randomisation will be achieved using minimisation including a random element that will allocate patients on a 1:1:1 basis to the three treatment strategies. If a patient is allocated to the alternative TNFi arm, then the patient will receive either etanercept (if patient failed to respond to monoclonal antibody) or a monoclonal antibody (if the patient failed to respond to etanercept). (See Figure 1) The trial is open-label, since it would be unreasonable to administer multiple dummy injections and

infusions as a means to keep the patient blinded to their true allocation. The primary endpoint is absolute reduction in DAS28 over a period of 24 weeks.

1.2.2 Early closure of trial

In November 2014, the HTA requested that the trial halt all further recruitment, and proceed with finishing follow-up for all randomised patients to a minimum of 48 weeks, and begin final analysis. The randomisation service closed to further recruitment with 122 patients randomised between 14th August 2012 and 18th December 2014.

1.3 Study aims and objectives

The study aims and objectives listed here are as provided in the protocol. Following the early closure of the study, the focus of the analysis will be on estimating the treatment effect of either experimental arm compared to rituximab in terms of disease activity; it is considered unlikely that a conclusion of non-inferiority (or superiority) will be reached.

1.3.1 Primary objective

To establish whether an alternative-mechanism-TNF-inhibitor (TNFi) or abatacept are non-inferior to rituximab in terms of disease response at 6 months (24 weeks) post randomisation.

1.3.2 Secondary objectives

- To compare alternative-mechanism-TNFi and abatacept to rituximab in terms of disease response over a 12 month period (48 weeks).
- To compare alternative-mechanism-TNFi and abatacept to rituximab in terms of quality of life, toxicity and safety over a 12 month period (48 weeks).
- To undertake an evaluation of the cost-effectiveness/health economics of switching patients to an alternative-mechanism TNFi, abatacept or rituximab.
- To compare structural and bone density outcomes for abatacept and alternative-mechanism TNFi to rituximab over a 12 month period (48 weeks), in terms of bone densitometry score.

1.3.3 Exploratory Objectives:

- To determine the optimal sequence of treatments by assessing whether the response to the second treatment in RA patients is affected by which of the initial TNFi groups the patients failed (anti-TNF monoclonal or TNF receptor fusion protein).
- To evaluate whether the response to the second treatment (alternative mechanism TNFi, abatacept or rituximab) is affected by whether the patient was a primary (no initial response) or secondary (loss of an initial response) response failure to their initial TNF-blocking therapy.
- To ascertain whether seropositive and seronegative (to rheumatoid factor +/-anti-cyclic-citrullinated peptide antibody) RA patients behave differently in their response and disease outcome measures to the three treatment arms, particularly with respect to rituximab

1.4 Sample size and expected accrual

1.4.1 Current Trial Design

A total of 477 participants were to be recruited.

Each experimental trial arm (alternative mechanism TNFi, abatacept) will be compared to rituximab for non-inferiority in terms of change in DAS28 at 6 months. In the following justification, no adjustment for multiplicity of the comparisons of each experimental trial arm to rituximab has been made. Each of the comparisons can be interpreted independently; the comparison between abatacept and rituximab will provide no information on the comparison between alternative mechanism anti-TNF and rituximab. Multiple comparison procedures are therefore not required when testing two independent hypotheses (37, 38).

A total of 429 evaluable participants are required to have 80% power for demonstrating non-inferiority of either abatacept or alternative mechanism TNFi to rituximab at the 5% significance level. A total of 143 evaluable participants in each treatment group will ensure that the lower limit of the two-sided 95% confidence interval for the true difference in DAS28 (abatacept/alternative mechanism TNFi – rituximab) lies above -0.6 units, assuming no difference between treatment groups and a standard deviation between participants of 1.8 units (REFLEX trial (3)). Allowing for a loss to follow-up of 10%, a total of 477 participants will be recruited.

The proposed non-inferiority margin of -0.6 units in the change in DAS28 at 6 months corresponds to the maximum difference in DAS28 score that is considered to be of no clinical relevance and is the threshold for the clinical distinction of 'inferiority' (corresponds to the maximum change in DAS28 in participants with low or moderate disease activity that is classified as "no response" by the EULAR criteria). DAS28 score of 0.6 units is also the reported measurement error (39).

For the secondary outcomes analysis to compare quality of life, toxicity and safety at 6 months between treatment arms our sample size of 143 evaluable participants per group would enable us to detect a standardised effect size of 0.33 (small to medium by the definition of Cohen (40)), with 80% power and a 2-sided 5% significance level.

1.4.2 Prior Trial Design

Prior to implementation of Protocol V6.0, the target recruitment was 870 patients. This would allow the trial to conclude that either arm were non-inferior to rituximab in terms of the proportions of patients achieving a DAS28 reduction of 1.2 or more without toxicity, and also detect a significant interaction effect between seropositivity status and treatment effect. For details as to the assumptions and original choice of non-inferiority margin, refer to the SWITCH protocol paper (4).

1.4.3 Planned Recruitment Rate

In order to recruit 477 participants the target recruitment rate was 0.5 to 0.75 patients per month per centre over a maximum of 40 sites across the UK, over a maximum of 53 months.

1.4.4 Final Recruitment

The SWITCH trial closed to further recruitment in December 2014 with 122 patients randomised.

1.5 Randomisation

Randomisation to one of the three study arms was performed centrally using the CTRU automated 24-hour telephone randomisation system. Authorisation codes, provided by the CTRU, were

required to access the randomisation system. These activities were performed by a member of the SWITCH research team.

Patients who gave written informed consent were first registered, and given a unique study ID number. Following completion of eligibility screening, patients who fulfilled the eligibility criteria were randomised to one of the three study arms.

Randomisations were achieved using minimisation incorporating a random element, via a computer program, that allocated patients in a 1:1:1 ratio between Alternative TNFi: Abatacept: Rituximab after taking account of the following factors, details of which will be required for randomisation:

- Randomising site
- Disease Duration (0 – 4 years, 5 or more years)
- Rheumatoid Factor / Anti-CCP status (Either seropositive, both seronegative)
- Pattern of TNFi non-response (Primary, Secondary)

After a randomisation is made to Alternative TNFi arm, the patient will be allocated to receive either Etanercept (if the previous TNFi failure was to a monoclonal antibody) or a monoclonal antibody (if the previous TNFi failure was to Etanercept). The treating clinician will choose the appropriate monoclonal antibody at his / her discretion.

In statistical analysis, underlined values will be taken as the reference category levels (estimating the effect of being eg Secondary Non-responder compared to Primary Non-Responder). Randomising Centre will not be fitted as a fixed effect in the analysis, so no reference category is required. See section 5.1.9, for how the random centre effect will be fitted.

1.6 Eligibility

Patients were required to satisfy the following criteria. Eligibility waivers to the inclusion / exclusion criteria were NOT permitted.

1.6.1 Inclusion Criteria

1. Male and female subjects aged ≥18 years at the time of signing the Informed Consent Form.

2. Patients with a diagnosis of rheumatoid arthritis as per the ACR/EULAR 2010 classification criteria confirmed at least 24 weeks prior to the screening visit.

3. Patients who have failed conventional DMARD therapy as per NICE/BSR Guidelines (41) i.e. failure of at least 2 DMARDS including MTX.

4. Patients with persistent RA disease activity despite having been treated with a current initial TNFi agent for at least 12 weeks. Active RA defined as*:

- Primary non-response: failing to improve DAS28 by > 1.2 or failing to achieve DAS28 ≤ 3.2 within the first 12 to 24 weeks of starting the initial TNFi. This may include patients that have shown a reduction in DAS28 of > 1.2 but still demonstrate unacceptably high disease activity in the physician's judgement with evidence of an overall DAS28 of ≥ 3.2

  OR
- Secondary non-response: defined as inefficacy to first TNFi (having demonstrated prior satisfactory response) as per clinician judgement; with intolerance not the reason for cessation of first TNFi.

  *These criteria are consistent with BSR guidelines (41).

5. MTX dose stable for 4 weeks prior to the screening visit and to be continued for the duration of the study.

6. Patients on NSAIDs and / or corticosteroids (oral prednisolone not exceeding 10mg daily) who have been on an unchanged regimen for at least 4 weeks prior to the screening visit and are expected to remain on a stable dose until the baseline assessments have been completed.

7. Provided written informed consent prior to any trial-specific procedures.

1.6.2 Exclusion Criteria

1.6.2.1 General

1. Major surgery (including joint surgery) within 8 weeks prior to the screening visit or planned major surgery within 52 weeks following randomization.

1.6.2.2 Study Specific

2. Patients with inflammatory joint disease of different origin, mixed connective tissue disease, Reiter's syndrome, psoriatic arthritis, systemic lupus erythematosus, or any arthritis with onset prior to 16 years of age.

3. Patients receiving doses of prednisolone > 10mg/day within the 4 weeks prior to the screening visit.

4. Patients receiving intra-articular or intra-muscular steroid injections within 4 weeks prior to the screening visit.

1.6.2.3 Excluded Previous or Concomitant Therapy:

5. Patients who have previously received more than 1 TNFi drug OR any other biological therapy for the treatment of RA.

6. Patients unable or unwilling to stop treatment with a prohibited DMARD (i.e synthetic DMARD aside from MTX e.g. oral or injectable gold, chloroquine, hydroxychloroquine, cyclosporine, azathioprine, leflunomide, sulphasalazine) prior to the start of protocol treatment.

7. Treatment with any investigational drug in the last 12 weeks prior the start of protocol treatment.

1.6.2.4 Exclusions for general safety

These criteria should be considered in the context of BSR guidance (41).

8. Patients with other co-morbidity including acute, severe infections, uncontrolled diabetes, uncontrolled hypertension, unstable ischaemic heart disease, moderate/severe heart failure (Class III/IV of the New York Heart Association (NYHA) functional classification system (42)), active bowel disease, active peptic ulcer disease, recent stroke (within 12 weeks before the screening visit), or any other condition which, in the opinion of the investigator, would put the patient at risk to participate in the study or would make implementation of the protocol difficult.

9. Patients with any major episode of infection requiring hospitalisation or treatment with IV antibiotics within 12 weeks of start of treatment protocol or oral antibiotics within 4 weeks of start of protocol treatment.

10. Patients at significant risk of infection, which in the opinion of the investigator would put the patient at risk to participate in the study (e.g. leg ulceration, indwelling urinary catheter, septic joint within 52 weeks (or ever if prosthetic joint still in situ)).

11. Patients with known active current or history of recurrent bacterial, viral, fungal, mycobacterial or other infections including herpes zoster (for tuberculosis and Hepatitis B and C see below), but excluding fungal infections of nail beds as per clinical judgment.

12. Patients with untreated active current or latent tuberculosis (TB). Patients should have been screened for latent TB (as per BSR guidelines) within 24 weeks prior to the screening visit and, if positive, treated following local practice guidelines prior to the start of protocol treatment.

13. Patients with active current hepatitis B and/or C infection. Patients should have been screened for hepatitis B and C within 24 weeks prior to the screening visit and if positive, excluded from the study.

14. Primary or secondary immunodeficiency (history of or currently active) unless related to primary disease under investigation.

15. Pregnancy, lactation or women of child-bearing potential (WCBP) unwilling to use an effective birth control measure whilst receiving treatment and after the last dose of protocol treatment as indicated in the relevant SmPC/IB.

16. Men whose partners are of child-bearing potential but who are unwilling to use an effective birth control measure whilst receiving treatment and after the last dose of protocol treatment as indicated in the relevant SmPC/IB.

1.6.2.5 Laboratory value exclusions

17. Patients with known significantly impaired bone marrow function as for example significant anaemia, leukopaenia, neutropaenia or thrombocytopaenia as shown by the following laboratory values at the time of the screening visit:

- Haemoglobin < 8.5 g/dl
- Platelet count < 100 x 10^9 / L
- White blood cell count < 2.0 x 10^9 / L

- Neutrophil count < 1 x 10^9 / L

18. Patients with known severe hypoproteinaemia at the time of the screening visit, e.g. in nephrotic syndrome or impaired renal function, as shown by:

- Serum Creatinine > 150 µmol / L

1.7 Planned analyses

No interim analyses will be undertaken prior to final analysis. One pre-specified interim analysis would have been conducted when 50% of patients had passed week 24, designed to allow for early stopping of an arm for demonstrating inferiority of either abatacept or alternative TNFi. With the closure of the trial before 25% of the expected patient numbers being recruited, the interim analysis is now obsolete.

The DMEC, in the light of the interim reports and of any advice or evidence requested, will if necessary report to the Trial Steering Committee (TSC) if there are concerns regarding the safety of the trial treatment.

2. Endpoints

The study endpoints are listed below. For definitions of endpoints (including references, where applicable) please see the endpoint definition sections 2.3 and 2.4.1 – 2.4.13.

2.1 Primary endpoint

The primary endpoint is the absolute change in DAS28 score (Disease Activity Score with 28 joint counts) between Baseline and Week 24.

2.2 Secondary endpoints

- DAS28 Score measured at Baseline, Week 12, Week 24, Week 36, Week 48.
- DAS28 “Response” at Week 12, Week 24, Week 36, Week 48.
- DAS28 Low Disease Activity at Baseline, Week 12, Week 24, Week 36, Week 48.
- DAS28 Remission at Baseline, Week 12, Week 24, Week 36, Week 48.
- EULAR Response Scores at Week 12, Week 24, Week 36, Week 48.

- EULAR / ACR Remission at Baseline, Week 12, Week 24, Week 36, Week 48.
- ACR Response Scores at Week 12, Week 24, Week 36, Week 48.
- CDAI Score at Baseline, Week 12, Week 24, Week 36, Week 48.
- SDAI Score at Baseline, Week 12, Week 24, Week 36, Week 48.

Quality of Life Endpoints

- RAQoL at Baseline, Week 12, Week 24, Week 36 and Week 48.
- HAQ-DI (also evaluated at weeks 60, 72, 84 and 96)
- HADS at Baseline, Week 12, Week 24, Week 36 and Week 48.
- Pain Visual Analogue Scale (also evaluated at weeks 60, 72, 84 and 96)
- General Health Visual Analogue Scale (also evaluated at weeks 60, 72, 84 and 96)
- Global Assessment of Arthritis Visual Analogue Scale (also evaluated at weeks 60, 72, 84 and 96)

Safety Endpoints (over 52 weeks)

- Toxicity
- Adverse Events

Economic Evaluation Endpoints

- EuroQol 5-dimensions (EQ-5DTM) (also evaluated at weeks 60, 72, 84 and 96)
- Health Utilities Index (also evaluated at weeks 60, 72, 84 and 96)
- Health and Social Care Use & Expenditure due to Rheumatoid Arthritis
- Incremental Cost Effectiveness

Imaging Endpoints

- Bone densitometry scan scores (T-scores unilateral neck of femur and lumbar spine-evaluated at baseline and week 48)

2.3 Primary endpoint definition

The DAS28 score to be used for the primary endpoint analysis is a composite measure of four items:

- Tender Joint Count (TJC: Range 0-28)
- Swollen Joint Count (SJC: Range 0-28)
- Erythrocyte Sedimentation Rate (ESR: Range 0-99)
- Patient-completed Visual Analogue Scale of Global Assessment of Arthritis, to answer the question "Considering all of the ways your arthritis has affected you, mark on the line below how you feel your arthritis is today" (VAS: Range "Very Well" = 0mm – "Very Poor" = 100mm)

With these four items, the DAS28 score is calculated in the following manner: (43,44)

$$DAS28 = \left(0.56 \times \sqrt{TJC}\right) + \left(0.28 \times \sqrt{SJC}\right) + (0.7 \times log_e ESR) + (0.014 \times VAS(mm))$$

Where LOG_e is the natural logarithm function, and $\sqrt{x}$ is the square root function.

Although other possible formulae exist for the DAS28 taking into account C-Reactive Protein (CRP) instead of ESR, or excluding ESR or CRP altogether, this is the definition of DAS28 that shall apply to the Primary Endpoint.

The Primary Endpoint is interpreted such that greater values indicate more active disease, and lower values indicate less active disease. Clinically relevant thresholds include Low Disease Activity (LDA) and Remission, both of which are defined in Section 2.4.2-2.4.3. EULAR disease response criteria consider the change from baseline as well as the present state, and are defined in Section 2.4.4.

For the Primary Endpoint Analysis, the absolute change from baseline in DAS28 shall be computed, by subtracting the follow-up value from the baseline value (see section 5.1.3). For this variable, positive values shall indicate worsening disease activity, and negative values shall indicate improving disease activity. Imputation of Missing Data items for the primary endpoint analysis is covered under section 2.6.1.

DAS28 can be categorised according to the value at a particular point in time as below:

Box 2 :DAS28 categories

High	5.1 < DAS28
Moderate	3.2 < DAS28 <= 5.1
Low	**2.6 < DAS28 <= 3.2**
Remission	**DAS28 <= 2.6**

Values in bold relate to key secondary endpoints at sections 2.4.2 and 2.4.3.

If any of the four components of DAS28 are missing, then the DAS28 value will be missing. See section 2.6 for how missing data will be imputed.

2.4 Secondary endpoint definitions

2.4.1 DAS28 "Response"

A patient will be deemed to have achieved a "Response" to treatment in terms of DAS28 (see section 2.3) if they achieve a reduction in DAS28 from baseline of 1.2 units or more. If the patient does has not achieved the required DAS28 reduction since baseline, the patient will be deemed to be a non-responder. If either the baseline or current values of DAS28 are not complete, then DAS28 "Response" will be missing. See section 2.6.1 for how missing data will be imputed.

2.4.2 DAS28 Low Disease Activity

A patient will be deemed to be in the state of Low Disease Activity (LDA) if at the assessment visit, their DAS28 score is in the interval (2.6, 3.2] (see section 2.3).

2.4.3 DAS28 Remission

A patient will be deemed to be in the state of Remission – both in terms of DAS28 (see section 2.3) and in terms of EULAR response – if at the assessment visit, their DAS28 score is 2.6 units or lower.

2.4.4 EULAR Response Criteria

European League Against Rheumatism (EULAR) Response criteria are determined according to the level of disease activity at the assessment, and by how much the DAS28 (see section 2.3) has improved since baseline. The diagram Box 3 illustrates how a patient is classified according to their Disease Activity and the improvement in disease activity.

A patient will be classed as having achieved No Response if:

- The DAS28 has reduced by less than 0.6 units (or has increased) since baseline, OR
- The DAS28 has reduced by between 0.6 and 1.2 units, and current DAS28 score is greater than 5.1 units.

A patient will be classed as having achieved Moderate Response if:

- The DAS28 has reduced by between 0.6 and 1.2 units, and current DAS28 score is 5.1 units or lower, OR
- The DAS28 has reduced by more than 1.2 units, and current DAS28 score is greater than 3.2 units.

A patient will be classed as having achieved Good Response if:

- The DAS28 has reduced by more than 1.2 units AND the current DAS28 score is 3.2 units or lower.

If the current DAS28 value, or the baseline value of DAS28 are not known, then the EULAR response level will be missing. See section 2.6 for how missing data will be imputed.

Box 3: EULAR response categories

Current DAS28 at endpoint	DAS28 IMPROVEMENT SINCE BASELINE		
	>1.2	<= 1.2 and ≥ 0.6	< 0.6
DAS28 <= 3.2	**GOOD** Response		
3.2 < DAS28 <= 5.1	**MODERATE** Response		
DAS28 > 5.1			**NO** Response

2.4.5 ACR Response Criteria (ACR20 / ACR50 / ACR70)

The American College of Rheumatology (ACR) Response criteria are composite measures developed for rheumatoid arthritis. There are three criteria that can be achieved, referred to as ACR20, ACR50 and ACR70. To achieve an ACR20, participants must demonstrate a relative improvement (reduction) from baseline of at least 20% (or 50%/70% for ACR50/ACR70 respectively) in both tender and swollen joint counts and also a relative 20% (or 50%/70%) improvement in 3 out of 5 following criteria (45):

- Patient global health assessment of disease activity (measured by a Visual Analogue Scale (VAS))
- Physician global assessment of disease activity (Measured by a VAS)
- Patient assessment of pain (Measured by VAS)
- Patient assessment of physical function (Measured by HAQ-DI© questionnaire)
- Results of laboratory test for inflammatory marker (Either erythrocyte sedimentation rate (ESR) or C-Reactive Protein (CRP))

2.4.6 ACR/EULAR Boolean remission rates (46)

Boolean remission is defined as swollen joint count (SJC), tender joint count (TJC), VAS patient global assessment (VAS) and CRP all ≤1.

2.4.7 SDAI (Simplified Disease Activity Index) score (44, 47)

The components of SDAI are the number of tender joints (28 joint count), the number of swollen joints (28 joint count), the patient global disease activity (10cm VAS), the physician global disease assessment (10cm VAS) and CRP (mg/dl). Since the SWITCH study records Visual Analogue Scales in mm in a range of 0-100mm, the VAS scores will first be scaled by dividing by 10. (10 being the conversion factor between centimetres and millimetres) Similarly, the SWITCH study records CRP in mg/L, so this will first be converted by dividing by 10. The SDAI is defined as:

$$SDAI = TJC + SJC + \left(\frac{PtVAS(mm)}{10}\right) + \left(\frac{PhVAS(mm)}{10}\right) + \left(\frac{CRP(mg/L)}{10}\right)$$

Box 4: SDAI disease activity states

High Disease Activity	26 < SDAI
Moderate Disease Activity	11 < SDAI <= 26
Low Disease Activity	3.3 < SDAI <= 11
Remission	0 <= SDAI <= 3.3

2.4.8 CDAI (Clinical Disease Activity Index) score (44, 48)

The components of the CDAI are: the number of tender joints (28 joint count), the number of swollen joints (28 joint count), a Patient global assessment of arthritis (10 cm VAS) and physician global assessment of arthritis (10 cm VAS). These are added to provide an assessment of disease activity on a scale of 0-76. Since the SWITCH study records Visual Analogue Scale scores in mm in a range of 0-100mm, the scores will first be scaled by dividing by 10. The CDAI is defined as:

$$CDAI = SJC + TJC + \left(\frac{PtVAS(mm)}{10}\right) + \left(\frac{PhVAS(mm)}{10}\right)$$

Box 5: CDAI disease activity states

High Disease Activity	22 < CDAI

Moderate Disease Activity	10 < CDAI <= 22
Low Disease Activity	2.8 < CDAI <= 10
Remission	0 <= CDAI <= 2.8

2.4.9 RAQoL (Rheumatoid Arthritis Quality of Life questionnaire) (49)

The RAQoL in a questionnaire that comprises 30 yes/no questions to which the patient responds. Each "Yes" scores 1 point. A fully-completed questionnaire is scored by summing the values gained for each question and takes a value in the range 0-30. Guidance is provided to deal with cases where a questionnaire is not fully-completed.

A summary of the scoring methodology is given in Appendix C.

2.4.10 HADS (Hospital Anxiety and Depression Scale)

The HADS is a questionnaire that comprises 16 questions, each of which is answered by the patient and has 4 possible responses. The questions are scored to take values in the range of 0-3. Half of the question scores are then summed to produce an overall Anxiety scale, with the other half being summed to produce an overall Depression scale.

A summary of the scoring methodology for the HADS is given in Appendix E.

2.4.11 HAQ-DI (Health Assessment Questionnaire – Disability Index)

The HAQ-DI is a questionnaire that comprises 24 questions, each of which is answered by the patient and has 4 possible responses. Questions relate to how much difficulty is experienced in undertaking certain activities, and whether any help or modified devices are required in order to complete them. The overall score is obtained from the average of 8 possible domains, each of which can take a value in the range of 0-3.

The HAQ-DI is a component of the ACR Response score. The HAQ-DI will need to be scored for all participants at all timepoints in order to compute the ACR response scores. A summary of the HAQ-DI scoring methodology is given in Appendix D.

2.4.12 Economic Evaluation Endpoints

The EQ-5D, Health Utilities Index, Health and Social Care Use and Expenditure due to Rheumatoid Arthritis and the Incremental Cost Effectiveness are endpoints of interest to the Health Economics Analysis. A separate plan will be written for such analysis, and the endpoints discussed in that document.

2.4.13 Toxicity

Toxicity is defined as the occurrence of an adverse event (including a serious adverse event, serious adverse reaction, adverse reaction, or SUSAR) that leads to permanent cessation of treatment.

2.5 Missing data

Data management will focus on the consenting process, participant eligibility, safety, dates and assessment results that feed into the primary and key secondary endpoints. Attempts will be made to retrieve missing data on these areas via a thorough data cleaning process. Every effort will be made to obtain key data items, as specified in the key data items document, all key data will be 100% checked for quality and completeness by either the Data Management Assistant or Data Manager. See Data Monitoring, Section 4.1 for further details.

The levels of missing data and reasons for missingness will be investigated for the consenting process, participant eligibility, safety, dates and assessment results. The quantity of missing data will be monitored by treatment group, and a summary of the number of patients with missing primary endpoint data and the quantity of missing data by treatment group and centre will be reported.

2.6 Imputation of missing data

Imputation of missing data under a model-based framework is limited by the expected number of observations at each timepoint. In order to account for both the longitudinal nature of the study, and the composite endpoints, it would be preferable to impute missing DAS28 and ACR Response components at all timepoints from baseline to week 48, and perform all imputations separately by

arm. However, since SWITCH did not recruit sufficient patients, such an imputation approach would require more observed patients at the final timepoint in each arm than were actually recruited. Accordingly, in order to satisfactorily impute data for the primary and key secondary endpoint analyses, DAS28 components will be imputed in a different framework to ACR Response components.

The decision to perform two separate approaches for the endpoints was taken to comply with three principles:

1. It is preferable to impute missing data separately within each treatment arm, rather than fitting a linear treatment covariate as a predictor. The former approach allows for the possibility of differential treatment modification effects for different components at different timepoints, while the latter approach does not.
2. The imputation model should, at a minimum reflect the analysis model. Accordingly, point (1) is important for the key secondary analysis of DAS28 values over time.
3. It is preferable to impute components separately, and then re-derive the composite endpoint than to directly impute the missing composite value. The latter approach ignores any known values that might contribute to the composite endpoint value.

The method of multiple imputation by chained equations (fully conditional specification) will be used to impute missing data for DAS28 components and ACR Response (50).

In each imputation, the missing value will be imputed in a model that includes the minimisation factors (excluding centre, owing to the large number of small centres) and the other components that make up the composite endpoint. The number of imputed datasets will be determined at analysis time, in the following manner:

1. In each model, the analysis dataset will be split into 3 parts, one for each treatment arm.
2. The percentage of missing values for each component at all timepoints will be determined in each arm.
3. The largest percentage of missing component-timepoint variables will be used to determine the number of imputations for that model.
4. The relevant percentage will be rounded up to the nearest whole percentage point, and one imputed dataset will be created for each percentage point of missingness indicated.

Thus, if the worst-completed variables in the three arms saw missingness as shown in Table 36:

Table 1: Example extent of missingness

Worst-completed components	TNFi missingness	Abatacept missingness	Rituximab missingness
TJC @48weeks	20.5%	0.6%	29.8%
SJC @48weeks	20.4%	15.8%	29.5%
ESR @48weeks	**30.1%**	15.7%	1.0%

Then 31 fully-imputed datasets will be created, regardless of the missingness of this value in other arms (50).

Since the some of the components are unlikely to be normally distributed – even after transformation - and may even be discrete values (in the case of the joint counts), predictive mean matching will be used to impute the missing value following the imputation (one observation from the 3 closest values to the predicted value will be chosen). In balancing the risk of biased imputations (due to choosing from too many neighbouring observations) and unstable results (choosing from too few) we bear in mind that the imputations will effectively be performed in subsamples of 40-41 patients, rather than 122. Since small sample sizes bring a risk of sparse observed data points in the vicinity of the predicted mean value, we choose to sample from the 3 nearest observations, to reduce the chance of the selected observation being far from the predicted mean value (51).

The longitudinal nature of the DAS28 (and ACR response) data over 48 weeks poses challenges for missing data imputation. In order to allow for correlation between visits to be accounted for, the data will be restructured into a “flat-file” format. For DAS28, a patient will have 20 components to be imputed, rather than 4 at 5 timepoints (to week 48). For ACR Response, a patient will have 24 components to be imputed, rather than 8 at 3 timepoints (to week 24).

Once the missing data items have been multiply imputed, the DAS28 or ACR Response will be derived and the analysis performed on each multiply-imputed dataset. The resulting parameter

estimates will be combined using Rubin's Rules for Multiple Imputation. The resulting parameter estimates will form the primary endpoint analysis (50, 52).

Patterns of missing data will be explored between the treatment arms, and potential relations to baseline characteristics and timing of missing data will be explored.

2.6.1 Primary Endpoint Analysis Imputation – DAS28

For the primary endpoint analysis of DAS28 reduction at 24 weeks and the key secondary endpoint analysis of DAS28 over a 48 week period, the components of the DAS28 (Tender and Swollen Joint Counts, ESR and Patient global assessment of arthritis) will be imputed if they are missing at any timepoint from baseline up to Week 48.

2.6.2 Secondary Endpoint analysis Imputation – ACR Response

For the key secondary endpoint analysis (ACR20 Response at 24 weeks), the components of the ACR Response criteria (Tender and Swollen Joint Counts, ESR, CRP, Physician global assessment of arthritis, Patient Pain assessment and Patient global assessment of arthritis) will be imputed if they are missing at any timepoint from baseline up to Week 24.

2.6.3 Imputation models

For each Imputation "effort", Multiple Imputation will be performed separately for each of the three treatment groups in isolation, rather than for the whole dataset incorporating a treatment group term. Missing values will be imputed in time order, starting with baseline values, then those at week 12, week 24, week 36 and finally week 48 in that order (for imputing ACR Response, imputations will cease after Week 24 values have been imputed). Within each visit timepoint, the missing values will be imputed in order from those with least missing data to those with most missing data.

To impute missing values for the (up to) 8 partially missing values at each timepoint, the multiple imputation procedure shall be invoked once, to impute all missing values required for that endpoint. Although we acknowledge that including additional variables in the imputation model can result in better imputed values and may make the Missing at Random assumption more plausible, we recognise that the expected small size of the dataset means that it would be infeasible to extend our

imputation models beyond that required for analysis. At the very least, we expect that our imputation models will match the analysis models.

2.6.4 Sensitivity Analyses

We will investigate the sensitivity of the conclusions to the Missing At Random Assumption by carrying out alternative methods of imputing missing DAS28 components, or scores:

A complete-case analysis (CC) will be performed, in which all participants missing at least one DAS28 component at Baseline or Week 24 will be completely excluded from the analysis. Such an analysis is not compatible with the Intention-to-Treat Analysis, and assumes that data is missing completely at random (MCAR). Differential non-completion may therefore result in biased treatment effect estimates.

3. Populations

3.1 Intention-to-treat population

An intention-to-treat analysis will be the primary method for analysing and summarising the trial data. The intention-to-treat population is defined as all randomised patients, regardless of if they are ineligible, withdrawn, don't comply with the protocol, are lost to follow-up or don't receive any study treatment. Only patients who have withdrawn their consent for their data to be used in the study (ie they have requested that their data be destroyed) or for whom written informed consent has not been received, will not be included in this population. These patients will be analysed and summarised according to the treatment they were randomised to receive.

3.2 Per protocol population

In the per-protocol population, patients will be analysed according to the treatment received. The per-protocol population will exclude patients whose trial conduct constitutes a major protocol violation (see Appendix B). A list of such violations will be discussed and agreed by the Chief Investigator prior to analysis.

For non-inferiority analysis of the Primary Endpoint, a null hypothesis of inferiority rejected in the ITT analysis population must also be rejected in the Per-Protocol analysis population for the conclusion of non-inferiority to be held.

3.3 Safety population

In the safety population, all participants will be included and safety data will be analysed according to the actual treatment received. If the patient is withdrawn from the study prior to receiving first dose of IMP, or the patient does not receive any IMP prior to completing the study, then the patient will be placed in a "Not received IMP" group, separate to the other possible treatment arms.

3.4 Quality of life populations

A separate quality of life population will be formed for the analysis of each questionnaire. (RAQoL, HAQ-DI, HADS) Each population will comprise all patients who return an analysable baseline questionnaire, regardless of subsequent questionnaire completion.

3.5 Complete Case (CC) Analysis Population

The CC analysis population will include all participants with all DAS28 components recorded at baseline and Week 24. Any patient missing any component at either visit will be excluded from this analysis population.

4. Data Handling

Data will be monitored for quality and completeness by the CTRU in the following areas; consenting process, participant eligibility, safety, date consistency and assessment results. Missing data in these areas will be chased until it is received, confirmed as not available or the trial is at analysis. Any problems with data collection will be discussed at internal project team meetings and, if appropriate, external project team meetings. All efforts will be made to ensure that as much of the data is present as possible and that reasons are obtained when data is unobtainable.

The CTRU/Sponsor will reserve the right to intermittently conduct source data verification exercises on a sample of patients, which will be carried out by staff from the CTRU/Sponsor. Source data verification will involve direct access to patient notes at the participating hospital sites and the on-going central collection of copies of consent forms and other relevant investigation reports. A Trial Monitoring Plan has been developed which details the standard data and process monitoring performed for this trial being conducted by the CTRU.

An independent data monitoring and ethics committee (DMEC) reviewed the safety and ethics of the trial as described in Section 1.7. The DMEC, in the light of the interim reports and of any advice or evidence they wish to request, (including the extent to which treating clinicians / investigators are complying with the protocol) were able to - if necessary - report to the Trial Steering Committee (TSC) if there were concerns regarding the safety of the trial treatment.

The following were also to be examined continuously during the course of the trial:

- Consent
- Recruitment
- Randomisation
- Data quality/completeness (priority will be given to the key data items used to analyse the primary endpoint)
- Compliance with the protocol(e.g. eligibility, contraindicated medications)
- SAEs/SUSARs/Deaths/Pregnancies
- Withdrawals from the trial / losses to follow-up

4.2 Data validation

Data management will focus on the data associated with the consenting process, participant eligibility, safety, date consistency and assessment outcomes and this section refers to the cleaning of these items. The Data Management Assistant (DMA)/ Data Manager (DM) will carry out initial validation of the forms in accordance with the trial-specific Data Management Work Instructions. This will ensure that data is complete, consistent, and up-to-date. The Data Clarification Form (DCF) will be sent to sites to highlight missing data items and queries associated with data on CRFS that appears to have inaccurate/ inconsistent data recorded. Reasons should be obtained when data is unobtainable.

The database will validate most data in line with validation rules and highlight any issues that need further investigation i.e. with the site. Manual checks on all entered data will be performed prior to the validations being implemented. Data items collected relating to the safety and rights of individual patients are to be highlighted via priority validations and dealt with as a data management priority. Periodic batch validation will also be carried out to detect any data queries that may be missed if case record forms (CRFs) are entered in an order that does not allow real time validation checks to work.

A key data items list drawn up by the Trial Statistician that will include all data items that are required for the analysis of the primary endpoint. All key data items will be checked manually for completeness and accuracy by the DMA/TC, in addition to any automatic checks raised on the database. Data automatically generated through the 24-hour randomisation system will be checked by the Trial Statistician.

The Trial Statistician will also perform checks to identify any missing or inconsistent data and liaise with the Data Manager to resolve any queries.

The data will be validated and checked using SAS in the following steps:

- The data will be read into permanent SAS data sets.
- A random sample of 5 patients from each SAS dataset were checked against the data as seen on the database to ensure that the data transfer has been successful, until such time as the download process was accepted to be working. The names and contents of the variables can be found on the annotated final database specification reports in the Statistician's Trial File.

Data checks will include:-

- Eligibility checks
- Sequential dates
- Checks for unusual and outlying data
- Inconsistency in data between forms
- Checks for missing data (are there variables which are systematically missing/do specific variables have a large amount of missing data, particularly key outcome data)
- Other checks as deemed appropriate

Any inconsistent data will be noted and an e-mail sent to the data manager responsible for the study. A copy of this e-mail will be kept in the statistician's trial file. All queries will be resolved and the outcome documented.

5. Data Analysis

5.1 General Principles

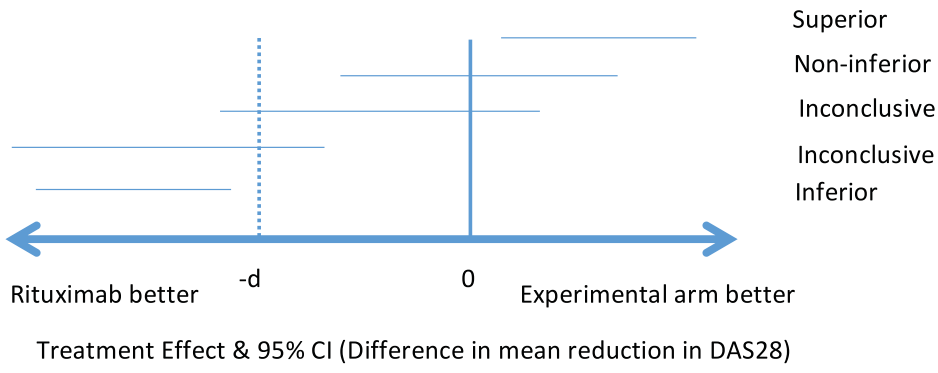
Unless otherwise stated, the Alternative TNFi arm shall be summarised as a single "Alternative TNFi" arm for summarising. Within this arm, listings will report either etanercept or the particular monoclonal antibody allocated (or received). The two comparisons of interest are between rituximab and Alternative TNFi, and between rituximab and abatacept.

All percentages will be calculated using the total number of patients within the specified analysis population as the denominator (i.e. including all patients with missing data for that variable), percentages will be reported to 1 decimal place. All statistical tests will be 2-sided and performed at the 5% significance level. All analyses will be carried out using SAS. Where all participants are included in categorical summaries, but percentages do not exactly sum to 100% due to rounding, a footnote will be included to the effect that percentages do not sum to 100% due to rounding.

5.1.1 Summary Statistics

Where "summary statistics" are requested of continuous-scale data, the number of non-missing items, the means, standard deviations, medians, upper and lower quartiles and minima and maxima will be summarised to one more decimal place than the data are collected. Values that are below the limit of detection and therefore non-quantifiable will be summarised using the limit of quantification value. For listings, if required, the non-quantifiable value would be reported as an inequality and the limit of quantification value used would also be reported. For categorical values, the number of values will be reported, along with the percentage of the whole population represented. Percentages will be reported to 1 decimal place.

Exploratory analyses will use informal hypothesis testing. All analyses will be carried out using SAS 9.4 unless otherwise stated.



-d: non-inferiority margin

Figure 2: Summary of interpretation of non-inferiority conclusions, as described in the CONSORT statement extension for non-inferiority studies (2).

5.1.2 Non-inferiority Analyses

The primary endpoint analysis of the SWITCH study is to be performed on a non-inferiority basis. The treatment effect will be estimated as described, and a 2-sided 95% confidence interval for the effect will be formed. If the lower bound of this confidence interval is above the pre-specified non-inferiority margin, then the treatment will be deemed to be non-inferior to rituximab.

The pre-specified non-inferiority margin for the primary endpoint is 0.6units of DAS28. If the 2-sided 95% confidence interval for the treatment effect is wholly below -0.6 units, then the conclusion of inferiority will be reached. If said confidence interval lies wholly above the margin, then the conclusion of non-inferiority will be reached.

Non-inferiority will be assessed in both the ITT population, and in the Per-protocol population. A conclusion of non-inferiority must be confirmed in both populations for the study to reach the overall conclusion that an experimental arm is not inferior to rituximab. If a 2-sided 95% confidence interval lies wholly above the null value of 0 for the intention to treat population, then it will be possible for the trial to conclude that an experimental arm is superior to rituximab. Figure 2

above illustrates the interpretation of the results, with regard to the upper and lower confidence limits.

5.1.3 Absolute changes from baseline

The absolute change from baseline will be computed as the value at follow-up minus the value at baseline. If either value is missing, then the absolute change from baseline will be missing. A brief interpretation of positive or negative change values will be provided.

For example: if a patient's swollen joint counts (SJC) are 10 at baseline, and 12 at follow-up, this will be a change of 2. If the SJCs are 12 at baseline and 10 at follow-up this will be a change of -2.

5.1.4 Relative changes from baseline

Relative changes from baseline are defined as the absolute change divided by the baseline value. If the absolute change is missing, then the relative change will also be missing. A brief interpretation of positive or negative change values will be provided.

For example: if a patient's RAQoL Scores are 8 at baseline, and 16 at follow-up, this will be a relative change of 1.0. If the RAQoL scores are 16 at baseline and 8 at follow-up this will be a relative change of -0.5.

Where a baseline value of 0 is recorded, a relative change from baseline will not be derived. However, in deriving the ACR Response categories, we will bear in mind that, for a baseline value of 0, there is no possible reduction that can yield either 20, 50 or 70% reduction. If a patient with a zero baseline component value records an increase (ie positive absolute change or deterioration) in this value, then we will impute this as a non-response for the relevant ACR component response criterion. However, if the absolute change at follow-up is zero, resulting in an undefined division of 0/0, then this patient's component response will be left as a missing value.

5.1.5 Percentage change

Within this document and the subsequent results, the phrase "Percentage Change" will be understood to be a *relative* change, as defined in Section 5.1.4. Accordingly, relative change values will be multiplied by 100 in order to express the relative changes as percentage changes.

Where a variable is to be measured on a scale of 0-100%, and the absolute change from baseline in this variable is required, then the unit of difference will be expressed as "percentage points" or "%age points", in order to differentiate from the phrase "Percentage Change", which is defined as the *relative* changes.

Thus, if a variable takes a baseline value of 50% and a follow-up value of 25%, this will be described as a 50% reduction (since the value has reduced by a half) or a reduction of 25%age points. Likewise, a variable with a baseline value of 20% and follow-up value of 100% will be described as an increase of 80 percentage points, but a 400% increase (since the follow-up value is four times greater than that at baseline).

5.1.6 Confidence Intervals for proportions

Confidence intervals for a single proportion shall be calculated using Exact Clopper-Pearson intervals. (Method 5 of (53)) Confidence intervals for an absolute difference between independent proportions shall be calculated using Exact intervals. (Method 8 of (54))

This will not apply to proportions estimated via logistic regression methods: fitted values for odds and 95% confidence intervals will be estimated, and these will then be back-transformed to the [0,1] probability scale.

This will also not apply when combining multiple-imputed datasets using Rubin's Rules. Instead, simple Wald-type confidence intervals will be used.

5.1.7 Randomisation errors

When handling the minimisation factors, patient data will be categorised as described for analysis. Where the data entered on the telephone randomisation system differs from any true values derived from baseline data, the corrected values resulting from the data cleaning process will be used for the primary analysis. Subgroup analyses will also use the corrected values. In addition to being the

principled approach, this will allow us to accommodate the change to the randomisation system where the balancing factor “RF status” was amended to “RF and ACPA status”.

5.1.8 Non-mutually exclusive selections

In summaries (for example of prior medical conditions, concomitant medication usage) where a single patient may reasonably have multiple responses selected, summaries will by default report only the values for each individual response level. No attempt will be made to enumerate a full list of all observed combinations unless specifically requested. Only particular pre-specified combinations of interest will be specified where appropriate. It will be assumed that a footnote to the effect of “These categories are not mutually exclusive” or “Patients may have multiple items selected” will be sufficient explanation for sums of percentages exceeding 100.

5.1.9 Multivariable modelling

Multivariable analyses will not be “built” following any model-fitting “strategy”. Instead, all variables specified for inclusion will be added to the model, and the significance of each factor will be reported. Where one categorical variable has more than one “factor level” then the significance of overall effect of including all factor levels will be tested, rather than those for each individual factor level. For all factor levels, suitable point and interval estimates of effect size will be presented.

Since we have two treatment comparisons of interest, our analysis will fit a single multivariable regression model, including the 3-level treatment variable. Then treatment contrasts will be formed, so as to compare the treatment effect of abatacept to rituximab, and to compare Alternative TNFi to rituximab.

Centre effects will be handled in accordance with Section 5.1.11.

5.1.10 Reference levels for categorical fixed effects

Where categorical variables are to be adjusted for in analyses, these shall use a pre-specified reference level. If the value is not pre-specified, then the modal value (ie the most frequently-

occurring) will be used. For the 3 minimisation factors that will be fixed effects, a pre-specified value will be used for the reference category (underlined in section S1.5).

For the treatment comparisons of interest, the reference category for the treatment effect will be rituximab.

5.1.11 Centre Effects

At the close of recruitment, 28 sites had randomised 122 patients. The median (and inter-quartile range) of by-centre recruitment was 3 patients (1-5), while 10 centres recruited between 1 and 2 patients in total. Owing to the large number of randomising centres with small numbers of patients, we will not attempt to fit a fixed effect for centre since a model fit is unlikely to converge: centre will be fitted as a random effect in the first instance. If an attempt to fit centre as a random effect fails to converge, then the centre will not be adjusted for in the analysis: centres will not be combined in any way so as to create a smaller number of larger pseudo-centres in order to allow the model fitting to converge and so randomising centre will be excluded from regression models. Where a decision is made to exclude a random centre effect from regression modelling, we will consider summaries that may support such a decision, including the intra-class correlation coefficient. (ICC)

5.1.12 Simulation and re-sampling methods

If any analysis requires the use of simulation and / or re-sampling methods, the initial "seed" value for the random number generation will be 20151902. The same seed will be used at the start of every such analysis.

5.1.13 Longitudinal Analyses

Analyses that model the effect of treatment over a period of time will be primarily be modelled as a random coefficients analysis as the primary analysis method, wherein the "time" effect will be directly calculated as the number of weeks since randomisation. A subsequent analysis for graphical purposes will use an alternative covariance-pattern model, in which the "time" effect is treated as a sequence of discrete timepoints, corresponding to the clinical assessment schedule. In

all such analyses, the baseline value will be fitted as a fixed effects covariate, rather than the first measurement at time t=0.

5.1.14 Visual Analogue Scales

Visual Analogue Scale (VAS) scores are measured on a scale of 0-100mm, and are usually only considered valid when scales of 100mm are used. Where sites have locally reproduced CRFs, rather than relying on professionally-printed CRF booklets, these scales will typically not be 100mm long. Rather than consider these scales to be missing data, we will rescale the VAS scores by dividing the position of the response by the measured line length, and multiplying the result by 100mm.

Table 2 below illustrates the outcome of this rescaling:

Table 2: Rescaling of Visual Analogue Scales

Line length	Position of response (from leftmost extremity of scale	Rescaled value
96mm	78mm	100 * 78 / 96 = 81.25mm
102mm	90mm	100 * 90 / 102 = 88.2…mm
94mm	(No mark)	Missing

5.2 Analysis

5.2.1 Baseline Characteristics

Summary statistics of baseline characteristics and pre-randomisation screening results will be presented by treatment arm and overall. Responses provided to questions during randomisation will be summarised. Where these differ from correct values provided on CRF, or derived values, these discrepancies will be listed.

5.2.2 Primary Endpoint Analysis

Primary Analysis

The observed DAS28 values at baseline, at week 24, and the absolute changes from baseline will be summarised by the three treatment arms. (Summary statistics are specified in S5.1.1)

The treatment effect of each experimental arm compared to rituximab will be estimated by means of a linear regression model, modelling the absolute change from baseline at Week 24 as a function of the experimental arm, the duration of arthritis category, the category of non-response and for the Rheumatoid Factor / Anti CCP seropositivity status. These variables will be included as fixed effects, and shall be categorised as described in Section 1.5 and Section 5.1.10. In the first instance, an attempt will be made to fit the randomising centre as a random effect, since most of the randomising centres are most likely too small for a fixed effect for centre to be successfully fitted. If this model does not converge, then centre will not be included in the regression model. As mentioned in Section 5.1.11, we will not combine small centres in any way to create a small number of larger pseudo-centres so as to improve the fit of the regression model.

As mentioned in Section 5.1.9, we will form treatment contrasts for the 3-level treatment group variable to compare the treatment effects of Abatacept vs Rituximab and the treatment effects of Alternative TNFi vs rituximab.

After fitting the model in each of the multiple-imputed datasets, and the resulting parameter estimates combined, the parameter estimate for each fixed effect will be presented along with its 95% confidence interval and the 2-sided P-Value under the hypothesis that the true parameter estimate is equal to zero.

The adequacy of the linear regression model for the primary analysis will be assessed by examining the following:

- Distribution of standardised residuals by predicted values;
- Adequacy of Normal distribution for residuals;
- Examining values of leverage to identify influential points;
- Correlation between residual values and order of enrolment.

Exploratory Analyses

There are 3 a-priori subgroup analyses planned. These are detailed in Section S5.3.1-S5.3.3. Within each treatment arm, patients will be subdivided as specified, and summary statistics reported within each subgroup.

5.2.2.1 Sensitivity Analyses

Complete Case Analysis

As detailed in Section 2.6.1, the primary endpoint will be analysed on a complete-case analysis basis: any participant missing at least one DAS28 component value at baseline or Week 24 will be excluded from the analysis.

5.2.3 Key Secondary Endpoint Analysis

Owing to the reduced level of recruitment and shortened trial timelines, a reduced amount of analysis will be conducted with respect to secondary endpoints.

5.2.3.1 DAS28 "Response" (reduction of 1.2 units or more) at 12, 24, 36, 48 weeks

The proportions of patients achieving this endpoint by arm at each timepoint will be summarised by treatment arm. (Summary statistics are specified in S5.1.1) After imputing missing values, the achievement of DAS28 "Response" will be analysed using a repeated measures random coefficients mixed effects logistic regression model, adjusting for the three minimisation factors (excluding centre) and baseline values of DAS28 (all modelled as fixed effects) and patient and patient by time effects (modelled as random effects) as well as time, randomised group and time by group interaction as fixed effects. Baseline values will be treated as a fixed effects covariate. It is not meaningful to include the baseline value as the first measurement at time t=0, since the DAS28 Response is based on change since baseline.

For graphical purposes, the mixed modelling analysis will also be performed using a covariance-pattern-type analysis, treating each visit as a sequence of discrete measurements, rather than a particular number of weeks.

5.2.3.2 DAS28 Score at 12, 24, 36, 48 weeks

Summary statistics of the DAS28 score will be presented at each timepoint by treatment arm. (Summary statistics are specified in S5.1.1) The DAS28 score will be analysed on a longitudinal analysis over the five visits from baseline to week 48. The values for the DAS28 components will be imputed as described, and the overall DAS28 score derived at each visit. Then, the values will be analysed using a random coefficients mixed effects linear regression model, adjusting for the three minimisation factors (excluding centre) and baseline value of DAS28 (all modelled as fixed effects) and patient and patient by time effects (modelled as random effects) as well as time, randomised group and time by group interaction as fixed effects. The baseline value will be treated as a fixed effect covariate, rather than the first measurement at time t=0. If attempts to fit a random effect for centre were not successful for the primary endpoint, then no attempt will be made to fit a random centre effect for this analysis.

For graphical purposes, the mixed modelling will be repeated as a covariance pattern-type analysis, treating the visits as separate discrete timepoints, rather than a number of weeks. Again, the baseline value will be treated as a fixed effects covariate, rather than the first measurement at time t=0.

5.2.3.3 ACR20 Response at Week 24

The proportions of participants achieving 20% reduction from baseline at week 24 in each of the ACR criteria will be summarised. (Summary statistics are specified in S5.1.1)

The binary variable ACR20 response at Week 24 will be analysed using a binary logistic regression model, adjusting for the 3 minimisation factors (excluding centre) all as fixed effects. If attempts to fit a random effect for centre were not successful for the primary endpoint, then no attempt will be made to fit a random centre effect for this analysis. If an attempt is made to fit a random centre effect in this analysis, and this is unsuccessful, the centre effect will not be included in the analysis. (See S5.1.11)

Once the model is fitted in each of the multiple-imputed datasets, and the resulting parameter estimates combined, the combined estimate of the odds ratio will be presented, along with its 95% confidence interval, and the 2-sided P-Value under the hypothesis that the Odds Ratio is 1.

5.2.4 Additional Secondary Endpoint Analyses

5.2.4.1 HAQ-DI

After scoring the HAQ-DI for all patients at all timepoints, summary statistics of the HAQ-DI at each timepoint will be presented by treatment arm and overall. (Summary statistics are specified in S5.1.1)

There will be no formal statistical analysis of this endpoint at any timepoint, as per the protocol Early Trial Termination Plan.

5.2.4.2 EULAR Response Scores

The frequency and proportions of participants achieving each level of EULAR response (no, moderate, good) at each timepoint (weeks 12, 24, 36, 48) participants will now be summarised by treatment arm and overall. (Summary statistics are specified in S5.1.1)

There will be no formal statistical analysis of this endpoint at any timepoint, as per the protocol Early Trial Termination Plan.

5.2.4.3 DAS28 Low Disease Activity and Remission states.

The frequency and proportions of participants achieving DAS28 Low Disease Activity and / or DAS28 Remission at each timepoint will be summarised by treatment arm and overall. (Summary statistics are specified in S5.1.1)

There will be no formal statistical analysis of this endpoint at any timepoint, as per the protocol Early Trial Termination Plan.

5.2.4.4 EULAR / ACR Remission

The frequency and proportions of participants achieving the EULAR / ACR Remission criteria at each timepoint will be summarised by treatment arm and overall. (Summary statistics are specified in S5.1.1)

5.2.4.5 ACR20, ACR50, ACR70 at Week 12, Week 24, Week 36 and Week 48

ACR20 at 24 weeks is already covered under 5.2.3.3. For all other response criteria at all other timepoints, the frequency and proportions of patients who achieve the particular response level at each timepoint will be summarised by treatment group and overall. (Summary statistics are specified in S5.1.1)

With the exception of the analysis planned for the ACR20 at Week 24, (as described in S5.2.3.3) there will be no formal statistical analysis of this endpoint at any timepoint, as per the protocol Early Trial Termination Plan.

5.2.4.6 Simplified Disease Activity Score

Summary statistics of the SDAI score at all timepoints will be presented by treatment arm and overall. The frequency and proportions of participants in each category of SDAI score will be summarised by treatment arm and overall. (Summary statistics are specified in S5.1.1)

There will be no formal statistical analysis of this endpoint at any timepoint, as per the protocol Early Trial Termination Plan.

5.2.4.7 Clinical Disease Activity Score

Summary statistics of the CDAI score at all timepoints will be presented by treatment arm and overall. The frequency and proportions of participants in each category of CDAI score at each timepoint will be summarised by treatment arm and overall. (Summary statistics are specified in S5.1.1)

There will be no formal statistical analysis of this endpoint at any timepoint, as per the protocol Early Trial Termination Plan.

5.2.4.8 RAQoL

It is recommended by the developers of the RAQoL that the score is treated only on an ordinal scale, and that summaries of the values are restricted to non-parametric statistics such as median, quartiles, minima and maxima.

Once scored, values of RAQoL will be summarised using non-parametric percentile-based summary statistics in each treatment group at each timepoint.

There will be no formal statistical analysis of this endpoint at any timepoint, as per the protocol Early Trial Termination Plan.

5.2.4.9 HADS

The HADS will be scored at each timepoint for all participants, providing the Anxiety and Depression scales. Summary statistics of both the Anxiety and Depression scores will be presented by treatment arm and overall at each timepoint. (Summary statistics are specified in S5.1.1)

There will be no formal statistical analysis of this endpoint at any timepoint, as per the protocol Early Trial Termination Plan.

5.2.4.10 Toxicity

The number of participants experiencing an adverse event leading to the permanent cessation of treatment will be summarised by arm. Within each arm, the timepoint at which treatment was permanently ceased will be summarised. Summaries will be presented based on both the ITT and Safety analysis populations. (Summary statistics are specified in S5.1.1)

5.2.4.11 Safety – AEs / SAEs / SARs / SUSARs / Deaths / Pregnancies

Summaries of Safety Data will be performed on the Safety population. (See section 3.3)

Numbers of Adverse Events, Serious Adverse Events, Serious Adverse Reactions, Suspected Unexpected Serious Adverse Reactions, Deaths and Pregnancies will be summarised by each arm, with numbers of participants experiencing at least one such event. Line listings of SAEs, SARs and SUSARs will be presented. Line listings of reported deaths and Pregnancies will be presented. For

Adverse Events, summaries of the suspected causalities, intensities and outcome / subsequent cessation of treatment will be provided.

5.2.4.12 Bone Densitometry

Summary statistics of t-scores and z-scores for spine and neck of femur will be presented at Baseline and at Week 48. (Summary statistics are specified in S5.1.1)

5.3 Subgroup Analyses

There are three a priori subgroup analyses planned, to investigate the possibility of a treatment modification effect on the primary endpoint. Any additional subgroup analyses will be deemed to be exploratory, and shall be described as such. Considering the small size of the study, with around 40 patients expected in each of the three arms, it is highly unlikely that any subgroup analyses will have sufficient power to make definitive conclusions as to any treatment modifying effect. No formal statistical analysis of subgroups will be performed. Owing to the reduced level of final recruitment, the amount of statistical analysis has been reduced to summary statistics. (Summary statistics are specified in S5.1.1)

5.3.1 Modification effect of initial TNFi failure on treatment effect

Summary statistics of change in DAS28 at 24 Weeks will be presented by treatment arm and overall. Within each treatment arm, the summaries will be presented by initial TNFi type.

5.3.2 Modification effect of response failure type on treatment effect

Summary statistics of change in DAS28 at 24 Weeks will be presented by treatment arm and overall. Within each treatment arm, the summaries will be presented by primary or secondary non-responder status.

5.3.3 Modification effect of Rheumatoid Factor (RF) / anti-cyclic-citrullinated peptide antibody (ACPA) seropositivity status on treatment effect

Summary statistics of change in DAS28 at 24 Weeks will be presented by treatment arm and overall. Within each treatment arm, the summaries will be presented in two categories: being either RF / ACPA seropositive or being both RF / ACPA seronegative.

5.4 Additional Patient Summaries and Analyses

Patient flow

In line with the CONSORT guidelines for reporting randomised controlled trials (1) – including its extension to non-inferiority studies (2) – a flow diagram shall illustrate the flow of patients through the study, including the strategies to which they were assigned, the test strategies actually received and the subsequent management of the patients through to end of follow-up. The flow diagram will include the numbers of patients contributing to each analysis population.

The reasons for patients not being randomised in the study will be summarised. The dates on which the first and final patients were randomised will be reported, along with the date of final follow-up for the last patient.

Withdrawals and loss to follow-up

The number of patient and PI withdrawals/loss to follow-up and reasons for these withdrawals will be summarised.

Protocol violators/deviations

Protocol violations/deviations will be summarised overall, by treatment group and centre, including violations of eligibility criteria on entry into the trial, deviations from the treatment and assessment schedule.

5.5 Serious breaches of GCP

All serious and potential breaches of GCP that have occurred throughout the trial will be summarised by the Trial Co-ordinator and presented in the final report.

6. Reporting and Dissemination of the Results

The trial has been registered with ClinicalTrials.gov (NCT01295151)

6.1 Authorship and acknowledgement

The success of the study depends upon the collaboration of all participants. For this reason, credit for the main results will be given to all those who have collaborated in the study, through authorship and by contribution. Uniform requirements for authorship for manuscripts submitted to medical journals will guide authorship decisions. These state that authorship credit should be based only on substantial contribution to:

- conception and design, or acquisition of data, or analysis and interpretation of data
- drafting the article or revising it critically for important intellectual content
- final approval of the version to be published
- and that all these conditions must be met (www.icmje.org).

Main trial-related publication: The Chief Investigator, as having conceived the study, overseeing the study (unless any future change) with overall responsibility and being central in drafting the article and interpretation of data shall be first author on the main trial-related publication. Co-Applicants and senior CTRU staff that also satisfy the above requirements will be named as co-authors in any publication, which will be discussed amongst the Trial Management Group (TMG) members. In addition, all collaborators will be listed as contributors for the main study publication, giving details of their roles in planning, conducting and reporting the study.

Additional trial-related publication(s): Whilst the exact composition of the main publication remains to be determined, there may be opportunities to publish additional reports associated with the trial. The nature of authorship will be discussed for such reports individually with the TMG but

may take the form of a different lead (first) author with the Chief Investigator as senior author for example.

The SWITCH team should be acknowledged in all publications, as should NIHR HTA (as detailed in Section 6.4 below). Other key individuals will be included as authors or contributors as appropriate and at the discretion of the TMG. The Trial Steering Committee (TSC) will resolve any disputes relating to authorship.

The Chairs and Independent members of the TSC and Data Monitoring and Ethics Committee (DMEC) will be acknowledged, but will not qualify for full authorship, in order to maintain their independence. Bristol-Myers Squibb shall also be acknowledged for providing drug.

Relevant NIHR Clinical Research Networks' (e.g. CCRN) support should be acknowledged appropriately in trial publications.

6.2 Data release

To maintain the scientific integrity of the study, data will not be released prior to the first publication of the results of the primary endpoint analysis, either for study publication or oral presentation purposes, without the permission of the DMEC and the TSC. The TSC will agree a publication plan and must be consulted prior to release or publication of any study data.

Individual collaborators must not publish data concerning their participants, which is directly relevant to the questions posed in the study until the main results of the study have been published. Local collaborators may not have access to study data until after publication of the main study results.

6.3 Processes for the drafting, review and submission of abstracts and manuscripts

The Chief Investigator as first author of abstracts is responsible for circulating these to the other members of the TMG and the Sponsor for review at least 15 days prior to the deadline for submission.

The agreed first author of manuscripts is responsible for ensuring:

- timely circulation of all drafts to all co-authors during manuscript development and prior to submission
- timely (and appropriate) circulation of reviewers' comments to all co-authors
- incorporation of comments into subsequent drafts
- communication with the TSC (i.e. ensuring submission is in line with TSC publication plan, and ensuring TSC receive the final draft prior to submission)

The Chief Investigator as first author is responsible for submission of the publication and must keep the TMG and all authors informed of the abstract's or manuscript's status. The TSC will be kept informed of rejections and publications as these occur. On publication, the first author should send copies of the abstract or manuscript to the TSC, the TMG, the Sponsor and to all co-authors, and ensure communication with NIHR HTA programme as outlined below.

6.4 Funder's Requirements

All materials to be submitted for publication (written, audio/visual and electronic) will be prepared and submitted to the NIHR Co-ordinating Centre for HTA (NCCHTA) in accordance with the NIHR HTA programme's requirements at the time a publication is drafted. This applies to all publications regardless of whether or not the primary results have been published.

7. Statistical Analysis Plan References

1. MOHER, D., S. HOPEWELL, K.F. SCHULZ, V. MONTORI, P.C. GOTZSCHE, P.J. DEVEREAUX, D. ELBOURNE, M. EGGER and D.G. ALTMAN. Consort 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. Bmj-British Medical Journal, 2010, 340.

2. PIAGGIO, G., D.R. ELBOURNE, S.J. POCOCK, S.J.W. EVANS, D.G. ALTMAN and C. GRP. Reporting of noninferiority and equivalence randomized trials extension of the consort 2010 statement. Jama-Journal of the American Medical Association, 2012, 308(24), pp.2594-2604.

3. COHEN, S.B., P. EMERY, M.W. GREENWALD, M. DOUGADOS, R.A. FURIE, M.C. GENOVESE, E.C. KEYSTONE, J.E. LOVELESS, G.-R. BURMESTER, M.W. CRAVETS, E.W.

HESSEY, T. SHAW, M.C. TOTORITIS and R.T. GRP. Rituximab for rheumatoid arthritis refractory to anti-tumor necrosis factor therapy - results of a multicenter, randomized, double-blind, placebo-controlled, phase iii trial evaluating primary efficacy and safety at twenty-four weeks. Arthritis and Rheumatism, 2006, 54(9), pp.2793-2806.

4. NAVARRO-COY, N.C., S. BROWN, A. BOSWORTH, C.T. DAVIES, P. EMERY, C.C. EVERETT, C. FERNANDEZ, J.C. GRAY, S. HARTLEY, C. HULME, A.M. KEENAN, C. MCCABE, A. REDMOND, C. REYNOLDS, D. SCOTT, L.D. SHARPLES, S. PAVITT and M.H. BUCH. The 'switch' study protocol: A randomised-controlled trial of switching to an alternative tumour-necrosis factor (tnf)-inhibitor drug or abatacept or rituximab in patients with rheumatoid arthritis who have failed an initial tnf-inhibitor drug. Bmc Musculoskeletal Disorders, 2014, 15.

5. MARKENSON, J.A. Worldwide trends in the socioeconomic impact and long-term prognosis of rheumatoid arthritis. Seminars in Arthritis and Rheumatism, 1991, 21(2 SUPPL. 1), pp.4-12.

6. HARRIS, E.D. Mechanisms of disease - rheumatoid-arthritis - pathophysiology and implications for therapy. New England Journal of Medicine, 1990, 322(18), pp.1277-1289.

7. LEE, D.M. and M.E. WEINBLATT. Rheumatoid arthritis. Lancet, 2001, 358(9285), pp.903-911.

8. KEATING, N., P. OTFINOWSKI, C. WENGER, J. FAST and L. DERKSEN. Understanding the caring capacity of informal networks of frail seniors: A case for care networks. Ageing and Society, 2003, 23, pp.115-127.

9. BARRETT, E.M., D.G.I. SCOTT, N.J. WILES and D.P.M. SYMMONS. The impact of rheumatoid arthritis on employment status in the early years of disease: A uk community-based study. Rheumatology, 2000, 39(12), pp.1403-1409.

10. WOLFE, F. The epidemiology of drug-treatment failure in rheumatoid-arthritis. Baillieres Clinical Rheumatology, 1995, 9(4), pp.619-632.

11. BROOKS, P.M. Clinical management of rheumatoid-arthritis. Lancet, 1993, 341(8840), pp.286-290.

12. FELDMANN, M., F.M. BRENNAN and R.N. MAINI. Role of cytokines in rheumatoid arthritis. Annual Review of Immunology, 1996, 14, pp.397-440.

13. ELLIOTT, M.J., R.N. MAINI, M. FELDMANN, A. LONGFOX, P. CHARLES, P. KATSIKIS, F.M. BRENNAN, J. WALKER, H. BIJL, J. GHRAYEB and J.N. WOODY. Treatment of rheumatoid-arthritis with chimeric monoclonal-antibodies to tumor-necrosis-factor-alpha. Arthritis and Rheumatism, 1993, 36(12), pp.1681-1690.

14. CHEN, Y.F., P. JOBANPUTRA, P. BARTON, S. JOWETT, S. BRYAN, W. CLARK, A. FRY-SMITH and A. BURLS. A systematic review of the effectiveness of adalimumab, etanercept and infliximab for the treatment of rheumatoid arthritis in adults and an economic evaluation of their cost-effectiveness. Health technology assessment (Winchester, England), 2006, 10(42), pp.iii-iv, xi-xiii, 1-229.

15. LIPSKY, P.E., D. VAN DER HEIJDE, E.W. ST CLAIR, D.E. FURST, F.C. BREEDVELD, J.R. KALDEN, J.S. SMOLEN, M. WEISMAN, P. EMERY, M. FELDMANN, G.R. HARRIMAN, R.N. MAINI and T. ANTI-TUMOR NECROSIS FACTOR. Infliximab and methotrexate in the treatment of rheumatoid arthritis. New England Journal of Medicine, 2000, 343(22), pp.1594-1602.

16. BUCH, M.H., S.J. BINGHAM, V. BEJARANO, D. BRYER, J. WHITE, P. EMERY, R. REECE and M. QUINN. Therapy of patients with rheumatoid arthritis: Outcome of infliximab failures switched to etanercept. Arthritis & Rheumatism-Arthritis Care & Research, 2007, 57(3), pp.448-453.

17. BUCH, M.H. and P. EMERY. Nonresponse to tumor necrosis factor antagonists - is there any point in re-treatment? Nature Clinical Practice Rheumatology, 2006, 2(6), pp.288-289.

18. BUCH, M.H., Y. SETO, S.J. BINGHAM, V. BEJARANO, D. BRYER, J. WHITE and P. EMERY. C-reactive protein as a predictor of infliximab treatment outcome in patients with rheumatoid arthritis - defining subtypes of nonresponse and subsequent response to etanercept. Arthritis and Rheumatism, 2005, 52(1), pp.42-48.

19. GOMEZ-REINO, J.J., L. CARMONA and B. GRP. Switching tnf antagonists in patients with chronic arthritis: An observational study of 488 patients over a four-year period. Arthritis Research & Therapy, 2006, 8(1).

20. HANSEN, K.E., J.P. HILDEBRAND, M.C. GENOVESE, J.J. CUSH, S. PATEL, D.A. COOLEY, S.B. COHEN, R.E. GANGNON and M.H. SCHIFF. The efficacy of switching from etanercept to infliximab in patients with rheumatoid arthritis. Journal of Rheumatology, 2004, 31(6), pp.1098-1102.

21. HJARDEM, E., M. OSTERGAARD, J. PODENPHANT, U. TARP, L.S. ANDERSEN, J. BING, E. PEEN, H.M. LINDEGAARD, V.S. RINGSDAL, A. RODGAARD, J. SKOT, A. HANSEN, H.H. MOGENSEN, J. UNKERSKOV and M.L. HETLAND. Do rheumatoid arthritis patients in clinical practice benefit from switching from infliximab to a second tumor necrosis factor alpha inhibitor? Annals of the Rheumatic Diseases, 2007, 66(9), pp.1184-1189.

22. HYRICH, K.L., M. LUNT, K.D. WATSON, D.P.M. SYMMONS, A.J. SILMAN and B.R. BRITISH SOC RHEUMATOL. Outcomes after switching from one anti-tumor necrosis factor alpha agent to a second anti-tumor necrosis factor alpha agent in patients with rheumatoid arthritis - results from a large uk national cohort study. Arthritis and Rheumatism, 2007, 56(1), pp.13-20.

23. NIKAS, S.N., P.V. VOULGARI, Y. ALAMANOS, C.G. PAPADOPOULOS, A.I. VENETSANOPOULOU, A.N. GEORGIADIS and A.A. DROSOS. Efficacy and safety of switching from infliximab to adalimumab: A comparative controlled study. Annals of the Rheumatic Diseases, 2006, 65(2), pp.257-260.

24. VAN VOLLENHOVEN, R., A. HARJU, S. BRANNEMARK and L. KLARESKOG. Treatment with infliximab (remicade) when etanercept (enbrel) has failed or vice versa: Data from the sture registry showing that switching tumour necrosis factor alpha blockers can make sense. Annals of the Rheumatic Diseases, 2003, 62(12), pp.1195-1198.

25. VAN VOLLENHOVEN, R.F. Switching between biological agents. Clinical and Experimental Rheumatology, 2004, 22(5), pp.S115-S121.

26. WICK, M.C., S. ERNESTAM, S. LINDBLAD, J. BRATT, L. KLARESKOG and R.F. VAN VOLLENHOVEN. Adalimumab (humira (r)) restores clinical response in patients with secondary loss of efficacy from infliximab (remicade (r)) or etanercept (enbrel (r)): Results from the sture registry at karolinska university hospital. Scandinavian Journal of Rheumatology, 2005, 34(5), pp.353-358.

27. SMOLEN, J.S., J. KAY, M.K. DOYLE and G.-A.S. INVESTIGATORS. Golimumab in patients with active rheumatoid arthritis after treatment with tumour necrosis factor a inhibitors (go-after study): A multicentre, randomised, double-blind, placebo-controlled, phase iii trial. (vol 374, pg 210, 2009). Lancet, 2009, 374(9699), pp.1422-1422.

28. GENOVESE, M.C., J. BECKER, M. SCHIFF, M. LUGGEN, Y. SHERRER, J. KREMER, C. BIRBARA, J. BOX, K. NATARAJAN, I. NUAMAH, T. LI, R. ARANDA, D.T. HAGERTY and M. DOUGADOS. Abatacept for rheumatoid arthritis refractory to tumor necrosis factor alpha inhibition. New England Journal of Medicine, 2005, 353(11), pp.1114-1123.

29. DASS, S., E.M. VITAL and P. EMERY. Development of psoriasis after b cell depletion with rituximab. Arthritis and Rheumatism, 2007, 56(8), pp.2715-2718.

30. EMERY, P., R. FLEISCHMANN, A. FILIPOWICZ-SOSNOWSKA, J. SCHECHTMAN, L. SZCZEPANSKI, A. KAVANAUGH, A.J. RACEWICZ, R.F. VAN VOLLENHOVEN, N.F. LI, S. AGARWAL, E.W. HESSEY, T.M. SHAW and D.S. GRP. The efficacy and safety of rituximab in patients with active rheumatoid arthritis despite methotrexate treatment - results of a phase iib randomized, double-blind, placebo-controlled, dose-ranging trial. Arthritis and Rheumatism, 2006, 54(5), pp.1390-1400.

31. ISAACS J.D., O.E., TAK P.P., DEODHAR A., KEYSTONE E., EMERY P., YOCUM D., HESSEY E., READ S. Autoantibody-positive rheumatoid arthritis (ra) patients (pts) have enhanced clinical response to rituximab (rtx) when compared with seronegative patients. Ann Rheum Dis., 2009, 68(Suppl3), p.442.

32. FINCKH, A., A. CIUREA, L. BRULHART, D. KYBURZ, B. MOELLER, S. DEHLER, S. REVAZ, J. DUDLER, C. GABAY and M. SWISS CLINICAL QUALITY. B cell depletion may be more effective than switching to an alternative anti-tumor necrosis factor agent in rheumatoid arthritis patients with inadequate response to anti-tumor necrosis factor agents. Arthritis and Rheumatism, 2007, 56(5), pp.1417-1423.

33. BUCH M., D.S., VITAL E., GOEB V., SZABO L., BRYER D. Is switching to rituximab more effective than switching to an alternative tumour necrosis factor blocking therapy (tnf-bt) in patients with rheumatoid arthritis (ra) who have failed previous tnf-bt? - single centre cohort experience. Ann Rheum Dis., 2009, 68(Suppl3), p.574.

34. CHATZIDIONYSIOU S., C.C., VAN VOLLENHOVEN R. Rituximab versus anti-tnf in patients who have previously failed one or more anti-tnfs in an observational cohort:The sarastra study. Ann Rheum Dis., 2009, 68(Suppl3), p.445.

35. NAM, J.L., K.L. WINTHROP, R.F. VAN VOLLENHOVEN, K. PAVELKA, G. VALESINI, E.M.A. HENSOR, G. WORTHY, R. LANDEWE, J.S. SMOLEN, P. EMERY and M.H. BUCH. Current evidence for the management of rheumatoid arthritis with biological disease-modifying antirheumatic drugs: A systematic literature review informing the eular recommendations for the management of ra. Annals of the Rheumatic Diseases, 2010, 69(6), pp.976-986.

36. SMOLEN, J.S., R. LANDEWE, F.C. BREEDVELD, M. DOUGADOS, P. EMERY, C. GAUJOUX-VIALA, S. GORTER, R. KNEVEL, J. NAM, M. SCHOELS, D. ALETAHA, M. BUCH, L. GOSSEC, T. HUIZINGA, J.W.J.W. BIJLSMA, G. BURMESTER, B. COMBE, M. CUTOLO, C. GABAY, J. GOMEZ-REINO, M. KOULOUMAS, T.K. KVIEN, E. MARTIN-MOLA, I. MCINNES, K. PAVELKA, P. VAN RIEL, M. SCHOLTE, D.L. SCOTT, T. SOKKA, G. VALESINI, R. VAN VOLLENHOVEN, K.L. WINTHROP, J. WONG, A. ZINK and D. VAN DER HEIJDE. Eular recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs. Annals of the Rheumatic Diseases, 2010, 69(6), pp.964-975.

37. O'KEEFE, D.J. Colloquy: Should familywise alpha be adjusted? Against familywise alpha adjustment. Human Communication Research, 2003, 29(3), pp.431-447.

38. PROSCHAN, M.A. and D.A. FOLLMANN. Multiple comparisons with control in a single experiment versus separate experiments - why do we feel differently. American Statistician, 1995, 49(2), pp.144-149.

39. VAN DER MAAS, A., E. LIE, R. CHRISTENSEN, E. CHOY, Y.A. DE MAN, P. VAN RIEL, T. WOODWORTH and A.A. DEN BROEDER. Construct and criterion validity of several proposed das28-based rheumatoid arthritis flare criteria: An omeract cohort validation study. Annals of the Rheumatic Diseases, 2013, 72(11), pp.1800-1805.

40. COHEN J. Statistical power analysis for the behavioral sciences 2nd ed. New Jersey: Lawrence Erlbaum Associates, 1988.

41. LEDINGHAM, J., C. DEIGHTON and SGAWG. Update on the british society for rheumatology guidelines for prescribing tnf alpha blockers in adults with rheumatoid arthritis (update of previous guidelines of april 2001). Rheumatology, 2005, 44(2), pp.157-163.

42. THE CRITERIA COMMITTEE OF THE NEW YORK HEART ASSOCIATION. Nomenclature and criteria for diagnosis of diseases of the heart and great vessels. 9th ed. Boston, Mass.: Little, Brown & Co, 1994.

43. PREVOO, M.L.L., M.A. VAN'T HOF, H.H. KUPER, M.A. VAN LEEUWEN, L.B.A. VAN DE PUTTE and P.L.C.M. VAN RIEL. Modified disease activity scores that include twenty-eight-joint counts development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. Arthritis & Rheumatism, 1995, 38(1), pp.44-48.

44. ANDERSON, J.K., L. ZIMMERMAN, L. CAPLAN and K. MICHAUD. Measures of rheumatoid arthritis disease activity patient (ptga) and provider (prga) global assessment of disease activity, disease activity score (das) and disease activity score with 28-joint counts (das28),

simplified disease activity index (sdai), clinical disease activity index (cdai), patient activity score (pas) and patient activity score-ii (pasii), routine assessment of patient index data (rapid), rheumatoid arthritis disease activity index (radai) and rheumatoid arthritis disease activity index-5 (radai-5), chronic arthritis systemic index (casi), patient-based disease activity score with esr (pdas1) and patient-based disease activity score without esr (pdas2), and mean overall index for rheumatoid arthritis (moi-ra). Arthritis Care & Research, 2011, 63, pp.S14-S36.

45. FELSON, D.T., J.J. ANDERSON, M. BOERS, C. BOMBARDIER, M. CHERNOFF, B. FRIED, D. FURST, C. GOLDSMITH, S. KIESZAK, R. LIGHTFOOT, H. PAULUS, P. TUGWELL, M. WEINBLATT, R. WIDMARK, H.J. WILLIAMS and F. WOLFE. The american-college-of-rheumatology preliminary core set of disease-activity measures for rheumatoid-arthritis clinical-trials. Arthritis and Rheumatism, 1993, 36(6), pp.729-740.

46. BYKERK, V.P. and E.M. MASSAROTTI. The new acr/eular remission criteria: Rationale for developing new criteria for remission. Rheumatology, 2012, 51(suppl 6), pp.vi16-vi20.

47. SMOLEN, J.S., F.C. BREEDVELD, M.H. SCHIFF, J.R. KALDEN, P. EMERY, G. EBERL, P.L. VAN RIEL and P. TUGWELL. A simplified disease activity index for rheumatoid arthritis for use in clinical practice. Rheumatology, 2003, 42(2), pp.244-257.

48. ALETAHA, D., V.P. NELL, T. STAMM, M. UFFMANN, S. PFLUGBEIL, K. MACHOLD and J.S. SMOLEN. Acute phase reactants add little to composite disease activity indices for rheumatoid arthritis: Validation of a clinical activity score. Arthritis Research & Therapy, 2005, 7(4), pp.R796-R806.

49. DE JONG, Z., D. VAN DER HEIJDE, S.P. MCKENNA and D. WHALLEY. The reliability and construct validity of the raqol: A rheumatoid arthritis-specific quality of life instrument. Rheumatology, 1997, 36(8), pp.878-883.

50. WHITE, I.R., P. ROYSTON and A.M. WOOD. Multiple imputation using chained equations: Issues and guidance for practice. Statistics in Medicine, 2011, 30(4), pp.377-399.

51. MORRIS, T.P., I.R. WHITE and P. ROYSTON. Tuning multiple imputation by predictive mean matching and local residual draws. BMC medical research methodology, 2014, 14(1), p.75.

52. RUBIN DB. Multiple imputation for nonresponse in surveys. New York: Wiley, 1987.

53. NEWCOMBE, R.G. Two-sided confidence intervals for the single proportion: Comparison of seven methods. Statistics in Medicine, 1998, 17(8), pp.857-872.

54. ---. Interval estimation for the difference between independent proportions: Comparison of eleven methods. Statistics in Medicine, 1998, 17(8), pp.873-890.

Appendix A: CONSORT Checklist for Non-inferiority Randomised Trials (Non-inferiority requirements in *italics*)

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title. *Identification as a noninferiority randomized trial in the title*	
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale. *Rationale for using a noninferiority design*	
	2b	Specific objectives or hypotheses. *Hypotheses concerning noninferiority, specifying the noninferiority margin with the rationale for its choice*	
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	
Participants	4a	Eligibility criteria for participants. *Whether participants in the noninferiority trial are similar to those in any trial(s) that established efficacy of the reference treatment.*	
	4b	Settings and locations where the data were collected	
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered. *Whether the reference treatment in the noninferiority trial is identical (or very similar) to that in any trial(s) that established efficacy*	
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed. *Specify the noninferiority outcome(s) and whether hypotheses for main and secondary outcome(s) are noninferiority or superiority. Whether the outcomes in the noninferiority trial are identical (or very similar) to those in any trial(s) that established efficacy of the reference treatment.*	
	6b	Any changes to trial outcomes after the trial commenced, with reasons	

Sample size	7a	How sample size was determined. *Whether the sample size was calculated using a noninferiority criterion and, if so, what the noninferiority margin was.*	
	7b	When applicable, explanation of any interim analyses and stopping guidelines. *To which outcome(s) they apply and whether related to a noninferiority hypothesis*	
Randomisation:			
Sequence generation	8a	Method used to generate the random allocation sequence	
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	
	11b	If relevant, description of the similarity of interventions	
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes. *Whether a 1- or 2-sided confidence interval approach was used*	
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	
	13b	For each group, losses and exclusions after randomisation, together with reasons	
Recruitment	14a	Dates defining the periods of recruitment and follow-up	
	14b	Why the trial ended or was stopped	

Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval). *For the outcome(s) for which noninferiority was hypothesized, a figure showing confidence intervals and the noninferiority margin may be useful.*	
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence. *Interpret results in relation to the noninferiority hypothesis. If a superiority conclusion is drawn for outcome(s) for which noninferiority was hypothesized, provide justification for switching*	
Other information			
Registration	23	Registration number and name of trial registry	
Protocol	24	Where the full trial protocol can be accessed, if available	
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	

Appendix B – Protocol Deviators to be excluded from the Per-Protocol Analysis

The Per-Protocol Population is a subset of all participants in the trial comprising those who are more compliant with the trial protocol, perhaps with respect to achieving a minimum level of exposure to trial medication, not having any major protocol deviations – including eligibility violations – and availability of assessment data.

Where a site / patient's conduct in the SWITCH trial constitutes a Major Protocol Deviation, the patient will be excluded from the Primary Endpoint Analysis in the per-protocol population.

The tables below list all major eligibility violations and protocol deviations, which will determine patients to be excluded from the Per-Protocol analysis population for the primary endpoint. In addition, a patient receiving a treatment other than that chosen at randomisation will be analysed in the treatment group corresponding to the treatment received.

Table B.1: Main categories of protocol deviation to be excluded from the Per-Protocol Population

Patient was found to be ineligible after randomisation for the reasons listed in Table B.2 (putting the patient either outside the target population, or having an impact on the outcome assessments)

Between Baseline and Week 24, the patient was not compliant with methotrexate (ie not between 80%-120% of required dose)

Not received at least 80% of expected infusions or injections of IMP, i.e. received fewer than:

- Two infusions of rituximab (Week 0 and Week 2)
- Four infusions of infliximab (out of Week 0, 2, 6, 14 and 22)
- Five injections of golimumab (out of Week 0, 4, 8, 12, 16 and 20)
- Ten (out of 12 fortnightly) injections of certolizumab pegol or adalimumab;
- Twenty (out of 24 weekly) injections of etanercept or abatacept

Week 24 assessment was more than 30 weeks from baseline visit

Participant has been over or under dosed – i.e. received a larger or smaller dose at one or more treatments.

Protocol treatment was interrupted for more than 28 days

Participant received additional treatment regarded as contraindicated (as stated in Section 12.7.1 of the protocol) between randomisation and end of protocol treatment (Week 48)

Participant received additional steroids at least 6 weeks prior to an endpoint related disease activity assessment (as stated in section 12.7 of the protocol)

Table B.2: Eligibility criteria which must be satisfied for membership of the Per-Protocol population

Inclusion Criteria

2. Patients with a diagnosis of rheumatoid arthritis as per the ACR/EULAR 2010 classification criteria confirmed at least 24 weeks prior to the screening visit.

3. Patients who have failed conventional DMARD therapy as per NICE/BSR Guidelines(41) i.e. failure of at least 2 DMARDS including MTX.

4. Patients with persistent RA disease activity despite having been treated with a current initial TNFi agent for at least 12 weeks. Active RA defined as:

- Primary non-response: failing to improve DAS28 by > 1.2 or failing to achieve DAS28 ≤ 3.2 within the first 12 to 24 weeks of starting the initial TNFi.
  This may include patients that have shown a reduction in DAS28 of > 1.2 but still demonstrate unacceptably high disease activity in the physician's judgement with evidence of an overall DAS28 of ≥ 3.2

OR

- Secondary non-response: defined as inefficacy to first TNFi (having demonstrated prior satisfactory response) as per clinician judgement; with intolerance not the reason for cessation of first TNFi.

5. MTX dose stable for 4 weeks prior to the screening visit and to be continued for the duration of the study.

6. Patients on NSAIDs and / or corticosteroids (oral prednisolone not exceeding 10mg daily) who have been on an unchanged regimen for at least 4 weeks prior to the screening visit and are expected to remain on a stable dose until the baseline assessments have been completed.

Exclusion Criteria

2. Patients with inflammatory joint disease of different origin, mixed connective tissue disease, Reiter's syndrome, psoriatic arthritis, systemic lupus erythematosus, or any arthritis with onset prior to 16 years of age.

3. Patients receiving doses of prednisolone > 10mg/day within the 4 weeks prior to the screening visit.

4. Patients receiving intra-articular or intra-muscular steroid injections within 4 weeks prior to the screening visit.

5. Patients who have previously received more than 1 TNFi drug OR any other biological therapy for the treatment of RA.

6. Patients unable or unwilling to stop treatment with a prohibited DMARD (i.e synthetic DMARD aside from MTX e.g. oral or injectable gold, chloroquine, hydroxychloroquine, cyclosporine, azathioprine, leflunomide, sulphasalazine) prior to the start of protocol treatment.

7. Treatment with any investigational drug in the last 12 weeks prior the start of protocol treatment.

Appendix C – Outline of scoring methodology for the RAQoL (Rheumatoid Arthritis Quality of Life Questionnaire)

The 30 questions of the RAQoL are scored to provide an overall single summary score. Each question is scored using a simple binary score: each "Yes" response scores 1 point, each "No"

scores 0 points. The points from each question are then summed to produce the overall RAQoL, as a score out of 30 points.

If a patient fails to respond to between 1 and 6 questions, then the score can still be computed as a score out of 30 by rescaling the total scored by the number of completed responses.

$$RAQoL = \begin{cases} S & if\ 0\ are\ missing \\ \dfrac{30}{30-m}S & if\ 1-6\ are\ missing \\ missing & if\ 7\ or\ more\ are\ missing \end{cases}$$

Where S is the sum of all points from the completed questions, and m is the number of missing items.

For example, if 26 questions are completed, and these scored a total of 19 / 26, the RAQoL score is 19 * (30/26) = 21.923 / 30.

It is recommended by the developers of the RAQoL that the score is treated only on an ordinal scale, and that summaries of the values are restricted to non-parametric statistics such as median, quartiles, minima and maxima.

Appendix D – Outline of scoring methodology for the HAQ-DI (Health Assessment Questionnaire – Disability Index)

The questions of the HAQ-DI are scored within 8 domains of activity which are then combined to provide an overall single summary score on a scale of 0-3. The score for each of the eight domains are derived from asking the respondent to report the level of difficulty experienced when undertaking certain activities, and whether any aids, devices or other help is required to complete these activities. The eight domains of activity are listed in column A of the table below.

(A) Domain	(B) Question Count	(C) Matching Aid / Device(s)
Dressing and Grooming	2	Devices used for dressing;
Arising	2	Special or built-up chair
Eating	3	Built-up or special utensils
Walking	2	Cane; Walker; Crutches; Wheelchair
Hygiene	3	Raised toilet seat; bathtub seat; bathtub bar; Long handled appliances in bathroom
Reach	2	Long-handled appliances for reach
Gripping and Opening	3	Jar opener (for jars previously opened)
Chores and Housework	3	-

Each question in each domain is scored between 0 and 3, with 0 corresponding to the least level of difficulty experienced (None at all) and 3 to the greatest level of difficulty experienced (Unable to do). Then, of the 2 or 3 questions in each domain (see column B for how many questions apply) the highest value is taken as the overall score for each domain. For example, if in the hygiene domain a patient has three responses scored 1, 1 and 3, then the overall score for the hygiene is 3, being the maximum value reported.

Once the score for each domain is determined, the score is then increased to account for any need to use aids and devices or help from others. If any of the matching aids or devices (Column C) are selected for that domain, or help is reportedly needed to undertake activities in this domain, then a domain score of 0 or 1 is increased to 2 (if the domain score is already a 2 or a 3, then this has no impact).

Finally, the overall score is determined by taking the average of all non-missing domain scores. If fewer than 6 domains have complete scores, then the HAQ score is missing.

Appendix E – Outline of scoring methodology for the HADS (Hospital Anxiety and Depression Scale)

The 14 questions of the HADS are grouped within 2 domains, which are scored and summarised separately. Question responses are scored with a value of 0-3, with 0 representing least level of anxiety or depression, and 3 representing the greatest level of anxiety or depression. After each question is scored (some questions require "reverse scoring", to account for responses being presented in different orders) the responses are summed within each domain to create the anxiety scale score and the depression scale score.

Anxiety Scale		**Depression Scale**	
Question	**Lowest Level**	**Question**	**Lowest Level**
Tense or "Wound Up"	0 = Not at all	Enjoy things	0 = Definitely as much
Frightened feeling	0 = Not at all	See funny side of things	0 = As much as always
Worrying thoughts	0 = Very little	Feel cheerful	0 = Most of the time
At ease and Relaxed	0 = Definitely	Slowed Down	0 = Not at all
Butterflies in stomach	0 = Not at all	Lost interest in appearance	0 = Just as much care as ever
Feel restless	0 = Not at all	Look forward to things	0 = As much as ever
Sudden panic	0 = Not at all	Enjoy good book etc	0 = Often

Approval of Analysis Plan

Clinical Trials Research Unit (CTRU)

The following analysis plan for the SWITCH study has been approved by the following personnel. Any signed amendments to the plan will be filed with this document.

Trial Statistician: Colin Everett

Signature: _____

Date: _____

Supervising Statistician: Sarah Brown

Signature: _____

Date: _____

Study Scientific Lead: Linda Sharples

Signature: _____

Date: _____

Senior Trial Co-ordinator: Claire Davies

Signature: _____

Date: _____

Data Manager: Catherine Reynolds

Signature: _____

Date: _____

CTRU Project Delivery Lead: Catherine Fernandez

Signature: _____

Date: _____

Chief Investigator: Maya Buch

Signature: _____

Date: _____

Additional information:

Analysis Plan Amendment and Deviation Log

Current Version Number	New Version Number *	Section	Description and Reason for Amendment or Deviation	Trial Statistician Name & Date	Supervising Statistician Name & Date
1.0		5.2.2 5.1.3	Planned primary endpoint analysis was to model change from baseline at Week 24 in DAS28, and consider the mean changes in experimental arms relative to a non-inferiority margin of -0.6 in light of the treatment contrast of Rituximab - experimental. Decision to reverse the comparison for presentation purposes, resulting in a "flipping" of the values about the line x=0 of the results figure, and consideration of the NI margin of 0.6. Achieved by quantifying "improvement" or "reduction" in DAS28, rather than "change". Original derived change values reversed by multiplying by -1, and repeating the analysis.	Colin Everett 01 SEP 2016	Sarah Brown 01 Sep 2016
1.0		S2.4.5	Planned key secondary endpoint analysis of ACR Response considered change in ESR or CRP equally: either a 20% reduction in ESR or a 20% reduction in CRP should count towards ACR20 Response at any timepoint for all patients. Following discussion as to whether the results of both blood tests should count at all follow-up times, or if e.g. the first instance of ESR 20% reduction should result in ignoring all subsequent changes in CRP (or vice versa) for that patient in determining ACR Response, a clinician informed both that the latter approach was correct. Following this log is a detailed description of how the endpoint was ultimately derived in SWITCH.	Colin Everett 01 SEP 2016	Sarah Brown 01 Sep 2016

* If the analysis plan is amended, note the new version number. If a deviation is made from the analysis plan within the analysis, leave blank.

1.0		S5.1.13	Planned analysis for longitudinal modelling was that random coefficient models would form the basis of quantifying the overall linear treatment effects over time, while covariance pattern models would be used to display graphically fitted values over time. Due to model assumptions for the random coefficients model not holding, covariance pattern models were fitted quantifying the varying the treatment effect over time.	Colin Everett 01 SEP 2016	Sarah Brown 01 Sep 2016

Implementation of the ACR Response criteria in SWITCH

The American College of Rheumatology Response criteria were interpreted to result in "consistency" of use of laboratory inflammatory markers to account for the longitudinal patient follow-up. The Table below illustrates the approach taken in SWITCH for five hypothetical patients when deciding whether ESR alone, CRP alone or both ESR and CRP should be used to derive ACR Response.

Patient	Baseline	Week 12	Week 24	Week 36	Week 48
A	ESR Known	**Response**	**Response**	**No Response**	**No Response**
	CRP Known	**Response**	Ignored (Non Resp)	Ignored (Resp)	Ignored (Resp)
B	ESR Known	**Did Not Attend**	**No Response**	**Response**	**No Response**
	CRP Known	**Did Not Attend**	**No Response**	Ignored (Non Resp)	Ignored (Resp)
C	ESR Known	No Response	Ignored (Resp)	Ignored (Resp)	Ignored (Resp)
	CRP Known	**Response**	***Missing***	**No Response**	**Response**
D	ESR Known	**No Response**	**No Response**	**No Response**	**Response**
	CRP Known	*Missing*	Ignored (Non Resp)	Ignored (Non Resp)	Ignored (Resp)
E	ESR Known	**Response**	**Missing**	**No Response**	**Response**
	CRP Missing	Unknown	Unknown	Unknown	Unknown

Table – Derivation of inflammatory marker improvement over length of follow-up. A bolded cell indicates that, at this follow-up visit, the value was used to determine if the inflammatory marker test result counted toward ACR Response. Cells in **bold** indicate which of the two test values are used in deriving ACR Response at each visit.

Patient A: Both ESR / CRP known at baseline. At Week 12, both ESR/CRP show (eg) 20% improvement. At week 24, ESR shows improvement, so CRP is disregarded at this visit, and all subsequent visits, even though doing so ignores required CRP response later on.

Patient B: As for patient A, except at Week 12 did not attend. At Week 24, ESR and CRP agree that no-response was seen. At Week 36, ESR Response is seen, so CRP is disregarded at that time, and for the final visit.

Patient C: At the first visit CRP response is seen, which overrules the ESR non-response at that visit. CRP is used from Week 24 onwards. At Week 24, no CRP value is available, so CRP response is missing. We do not use the known ESR response value at this visit: we continue using only CRP, even if it is not known at a visit.

Patient D had both values known at baseline, but the CRP value was missing at the first observed follow-up visit. Since the ESR value was known, the ESR value is used for all future visits, and CRP is ignored.

Patient E had a missing CRP at baseline. It is not possible to derive a relative change from baseline when the baseline value is not known, so the ESR value is used exclusively at all visits. Had the ESR value also been unknown, then neither ESR nor CRP response / non-response can be derived at any subsequent visit.