Author(s): Connock, M., Cummins, E., Sutcliffe, Paul A., Court, Rachel A., Johnson, Rhona, Grove, Amy L., Ngianga-Bakwin, Kandala, Mohiuddin, Syed and Clarke, Aileen

Article Title: Belimumab for the treatment of active autoantibody-positive systemic lupus erythematosus

Year of publication: 2011

Link to publication: http://guidance.nice.org.uk/TA/Wave25/12/Consultation/EvaluationReport/EvidenceReviewGroupReport

Evidence Review Group Report commissioned by the National Health Service Research & Development Programme on behalf of the National Institute for Health and Clinical Evidence

Title: Belimumab for the treatment of active autoantibody-positive systemic lupus erythematosus

Produced by: WARWICK EVIDENCE

Authors:
Martin Connock, Senior Research Fellow, Warwick Evidence
Ewen Cummins, Health Economist, McMDC Ltd.
Paul Sutcliffe, Senior Research Fellow, Warwick Evidence
Rachel Court, Information Specialist, Warwick Evidence
Rhona Johnson, Computer Programmer, McMDC Ltd
Amy Grove, Project Manager, Warwick Evidence
Kandala Ngianga-Bakwin, Principal Research Fellow, Warwick Evidence
Syed Mohiuddin, Research Fellow, Warwick Evidence
Aileen Clarke, Director for Warwick Evidence, Professor of Public Health & Health Services Research, Warwick Evidence

Correspondence to: *Dr Martin Connock, Warwick Evidence, Health Sciences Research Institute, Warwick Medical School, University of Warwick, Coventry CV4 7AL*

Date completed: *26.06.2011*

Source of funding:
This report was commissioned by the National Institute for Health Research Health Technology Assessment Programme as project number: 10/73/01

Declared conflicts of interest of authors:
None

Acknowledgements:
We gratefully acknowledge the expert clinical advice input from Professor C. Gordon (Professor of Rheumatology, University of Birmingham).

Responsibility for report:
The views expressed in this report are those of the authors and not necessarily those of the NIHR HTA Programme or the NICE.

This report should be referenced as follows:
Connock M, Cummins E, Sutcliffe P, Court R, Johnson R, Grove A, Ngianga-Bakwin K, Mohiuddin S, and Clarke A; Belimumab for the treatment of active autoantibody-positive systemic lupus erythematosus: A Single Technology Appraisal. Warwick Evidence, 2011.

Contributions of authors:
Connock, Sutcliffe, Clarke – critical appraisal of clinical effectiveness. Cummins, Johnson, Ngianga Bakqin, Mohiuddin – critical appraisal of economic modelling. Ngianga-Bakwin – statistical appraisal. Court – searches. All authors – writing the report.

Table of Contents

Table of Tables

Table of Figures

Table of abbreviations

AC	Appraisal Committee
ACR	American College of Rheumatology
AE	Adverse Event
AMS	Adjusted Mean SLEDAI
ANCOVA	Analysis of Covariance
BILAG	British Isles Lupus Assessment Group
BLISS	Belimumab International SLE Study
BLyS	B-Lymphocyte Stimulator
BRAM	Birmingham Rheumatoid Arthritis Model
C	Complement
C3	Complement Component 3
C4	Complement Component 4
CAPD	Cumulative Average Prednisone Dose
CEAC	Cost Effectiveness Acceptability Curve
CI	Confidence Interval
CRD	Centre for Reviews and Dissemination
CVD	Cardiovascular Disease
EMA	European Medicines Agency
EQ-5D	EuroQoL 5 dimensions
ERG	Evidence Review Group
FACIT	Functional Assessment of Chronic Illness Therapy
FAD	Final Appraisal Determination
FDA	Food and Drug Administration
HDAS	High Disease Activity Subgroup
HGS	Human Genome Sciences
HRG	Health Research Group
HRQoL	Health-Related Quality of Life
ICER	Incremental Cost-Effectiveness Ratio
ISRCTN	International Standard Randomised Controlled Trial Number
IV	Intravenous
JHU	Johns Hopkins University
MeSH	Medical Subject Headings
MS	Manufacturer's Submission
NHS	National Health Service
NICE	National Institute for Health and Clinical Excellence
NSAIDs	Non-Steroidal Anti-Inflammatory Drugs
OR	Odds Ratio
PAS	Patient Access Scheme
PCS	Physical Component Summary
PGA	Physicians Global Assessment
PSA	Probabilistic Sensitivity Analysis
PSS	Personal Social Services
PSSRU	Personal Social Services Research Unit
QALY	Quality Adjusted Life Year
QoL	Quality of Life
RCT	Randomised Controlled Trial
SELENA	Safety of Estrogen in Lupus National Assessment
SF-36	Short Form 36-Item Health Survey
SLE	Systemic Lupus Erythematosus

SLEDAI	Systemic Lupus Erythematosus Disease Activity Index
SLICC	Systemic Lupus International Collaborating Clinics
SoC	Standard of Care
SPC	Summary of Product Characteristics
SRI	SLE Responder Index
SS	SELENA-SLEDAI
STA	Single Technology Appraisal
UK	United Kingdom
USA	United States of America
URTI	Upper Respiratory Tract Infection
UTI	Urinary Tract Infection
VB	Visual Basic
Vs.	Versus
WTP	Willingness to Pay

Glossary of terms

POPULATION	SYNONYMS	DEFINITION	SOURCE
Auto-antibody positive active SLE	As BLISS trial populations	Active SLE disease, defined as a SELENA-SLEDAI (SELENA=Safety of Estrogens in Systemic Lupus Erythematosus National Assessment; SLEDAI=Systemic Lupus Erythematosus Disease Activity Index) score ≥6 and positive anti-nuclear antibody (ANA) test results (ANA titre ≥1:80 and/or a positive anti-dsDNA [≥30 units/ml]) at screening	Summary of Product Characteristics (Page 8) Manufacturer's clarification document (A1)
Proposed license population	Marketing authorisation population. High disease activity Population A	This subgroup comprises patients with positive anti-dsDNA and who also had low complement C3 or C4 Benlysta is indicated for reducing disease activity in adult patients with active, autoantibody-positive systemic lupus erythematosus (SLE) with a high degree of disease activity (e.g positive anti-dsDNA, low complement) despite standard therapy	Manufacturer's submission (Page 20) and clarification document (A1) Summary of Product Characteristics (Page 15)
Target population	High disease activity population that is the focus of submission	Adults with active autoantibody-positive systemic lupus erythematosus with evidence for serological disease activity (low complement, positive anti-dsDNA) and SELENA-SLEDAI ≥10)	Manufacturer's submission (Page 20)

1 SUMMARY

1.1 Scope of the manufacturer submission

The manufacturer's scope encompasses the clinical effectiveness and cost effectiveness of belimumab plus Standard of Care (SoC) relative to SoC alone, for the treatment of adults with active auto-antibody positive Systemic Lupus Erythematosus (SLE) and also for a subgroup of these patients who exhibit signs of high disease activity. According to the manufacturer's scope and submission the population of greatest interest is the sub-group with high disease activity called the Target population. No subgroup is specified in the National Institute for Health and Clinical Excellence (NICE) scope. The Target population is a subgroup of the proposed licensed population; a decision on the manufacturer's license application is awaited.

The manufacturer's scope specifies that belimumab is delivered at 10mg/kg via a 1 hour intravenous (IV) infusion at 2 week intervals for the first 3 doses and every 4 weeks thereafter. The NICE scope merely states that belimumab is used as an add-on to SoC. The manufacturer's scope specifies SoC as the sole comparator, and considers there is no credible evidence to enable a statistical comparison of belimumab with other drugs, either rituximab or cyclophosphamide. These additional comparators are however specified in the NICE scope.

Outcomes listed in the manufacturer's scope include: disease activity, incidence and severity of flares, mortality, health-related quality of life (HRQoL), adverse effects of treatment, and fatigue; these coincide with those in the NICE scope.

The manufacturer's scope for economic analysis specifies a lifetime horizon, a National Health Service (NHS) and Personal Social Services (PSS) perspective, and a cost effectiveness analysis expressed in terms of incremental cost per quality adjusted life year; this corresponds to the NICE scope.

Special considerations raised in the manufacturer's scope include: 1) the innovative nature of belimumab for SLE; 2) the inability of the utility method to sensitively capture the quality of life (QoL) of SLE patients; and 3) the impact of SLE on particular ethnic groups and on women of childbearing age. There were no equity issues identified in the NICE scope.

The Food and Drug Administration (FDA) granted marketing authorisation in the United States of America (USA) in March 2011. Belimumab does not yet have marketing authorisation in Europe, the decision on an application is pending. It is therefore not yet certain if the manufacturer's scope will conform to directives from the European Licensing Authority.

Phase III trials have examined the effectiveness of belimumab at dosage regimens of 1mg/kg and 10mg/kg. The evidence relating to the 1mg/kg dose regimen has not been presented in the submission.

1.2 Summary of clinical effectiveness evidence submitted by the manufacturer

A systematic literature search was undertaken by the manufacturer although the results of subsequent analysis of the retrieved studies were not clearly reported.

The submitted clinical effectiveness evidence is mainly derived from two international multicentre phase III randomised placebo controlled trials (Belimumab International SLE Study (BLISS)-52 and BLISS-76 lasting 52 and 76 weeks, respectively), which compared SoC plus belimumab vs. SoC plus placebo. In each trial patients were randomised (approximately 1:1:1) to one of three treatments: SoC + placebo, SoC + belimumab at 1mg/kg, or SoC + belimumab at 10mg/kg. BLISS-52 was undertaken mainly in Asia and South America while BLISS-76 patients mainly derived from North America and Europe.

There were 288 and 271 patients in the 1mg/kg arms of BLISS-52and BLISS-76 respectively, but results for effectiveness in these groups were not submitted. Of 865 patients in BLISS52 and 819 in BLISS-76, 577 and 548 were distributed almost equally into placebo and 10mg/kg belimumab groups, respectively.

Several sets of results were presented for placebo and 10mg/kg belimumab arms of the trials: [i] those for the whole population of trial patients, separately by trial and also after pooling patients from the two trials; [ii] those for the "Target population" after pooling data across the two trials; the Target population is a subgroup with high disease activity identified by post hoc analyses. The submission of multiple sets of results complicates any summary of the clinical effectiveness data.

For the Target population some outcome by-trial results became available during the clarification process, however the Evidence Review Group (ERG) was unable to comment on within trial comparison of belimumab vs. placebo for: the mean change in Short Form 36 Item Health Survey (SF-36) Physical Component Survey (PCS) score; time to first flare; SLICC/ACR (Systemic Lupus International Collaborating Clinics / American College of Rheumatology); and Functional Assessment of Chronic Illness Therapy (FACIT)-Fatigue Score. Box 1 provides a summary of the manufacturer's response and reasons for focussing on the Target population.

Box 1: Interpretation of marketing authorisation population and the Target population

Our submission is based on a high disease activity subgroup of the marketing authorisation population, defined as the Target population. We acknowledge that NICE will be unable to make a recommendation for the whole of the expected licensed population (marketing authorisation population), but are aware that our Target population falls within the expected licensed population. Mindful of the need to make the most efficient use of NHS resources, this subgroup allows a Targeted approach to selecting the patients who are most likely to get the greatest benefit from treatment.

1.2.1 Primary outcome

The BLISS trials employed a novel composite outcome measure called the SLE Responder Index (SRI) which aimed to detect any improvement in SLE manifestation while guarding against the possibility that worsening involvement of other organ systems or a worsening in overall disease activity might be masked. To be classified as a "responder" a patient was required to satisfy specified minimum criteria in three measures of change in disease activity relative to baseline. The measures used were: [1] the Safety of Estrogen in Lupus Erythematosus National Assessment-SLE Disease Activity Index (SELENA SLEDAI) score, which detects an improvement in SLE manifestations, it scores disease activity over a range of 0 to 105 points and encompasses 24 weighted items scored as present or absent in the previous 10 days; clinically meaningful differences have been reported to be an improvement by a decrease of 6 points or a worsening by an increase of 8 points[1]; [2] the British Isles Lupus Activity Group (BILAG) index assesses organ system involvement over the preceding 28 days and is capable of detecting worsening of organ system involvement; it includes 86 items grouped into 8 organ systems, general (5 items), mucocutaneous (18 items), neurological (15 items), musculoskeletal (9 items), cardiorespiratory (12 items), vasculitis (8 items), renal (11 items), and hematological (8 items), (Isenberg and Gordon, 2000).[2] A score is calculated for each system depending on the SLE clinical manifestations (or signs and symptoms) present and whether they are new, worse, the same, improving, or not present in the last 4 weeks compared with the previous 4 weeks. BILAG uses classifications ranging from A to E as follows: A = worsening judged to require intensification of steroids or immunosuppressant treatments; B = worsening judged to require antimalarials, nonsteroidal anti-inflammatory drugs (NSAIDs), or low dose steroids; C = stable disease; D = improvement; E = system never involved; [3] a Physicians Global Assessment (PGA) score, employed to monitor for worsening in patient overall disease activity (scores can range between no disease = 0, and 3 = severe disease). The SRI criteria used to define a responder

were: an improved SLEDAI score by ≥ 4 points; a BILAG index showing no new grade A organ involvement or no two grade B organ involvements; a PGA score that has not increased by more than 0.3 points. The primary end point in both trials was the proportion of responders at week 52 relative to baseline according to the SRI.

In summary, the primary efficacy endpoint was the response rate at week 52, assessed with SLE Responder Index (SRI). A responder was defined as having a reduction of at least 4 points in the SELENA-SLEDAI score, no new BILAG A organ domain score, no more than 1 new BILAG B organ domain score, and no worsening in PGA score (increase < 0.3) at week 52 compared with baseline.

In both trials at 52 weeks SoC + 10mg/kg belimumab delivered a greater percentage of responders than did SoC + placebo. The difference in percentage of responders in the belimumab group relative to placebo group was 14% in BLISS-52and 9.4% in BLISS-76 and 11.8% for the whole population pooled across trials. The corresponding adjusted odds ratios for a response in BLISS-52 and in BLISS-76 were respectively 1.83 (95% CI: 1.30, 2.59; P = 0.0006) and 1.52 (95% CI: 1.07, 2.15; P = 0.027).

For the high disease activity subgroup (Target population) pooled across trials the difference in percentage of responders between the belimumab group and placebo group was 24.8% and the adjusted odds ratio was 2.7 (95% CI: 1.8, 4.1; P < 0.0001). For the Target population in BLISS-52 the difference in percentage of responders between the belimumab group and placebo group was 25.9% and the adjusted odds ratio was 3.0 (95% CI: 1.7, 5.2; P < 0.0001). For the Target population in BLISS-76 the difference in percentage of responders between the belimumab group and placebo group was 22.4% and the adjusted odds ratio was 2.5 (95% CI: 1.3, 4.6; P < 0.0045).

The percentage of responders was also reported at multiple follow up times. For the Target population pooled across trials and in BLISS-52, at many times, a significantly greater response was observed for the belimumab group relative to placebo group (significance tests uncorrected for multiple testing), however for BLISS-76 the only time a significantly (P < 0.05) greater response was observed for the belimumab group was at week 52.

1.2.2 Secondary Outcomes

The pre-specified major secondary outcomes were: the SRI response at week 76; the percentage of patients with a ≥ 4 point SLEDAI improvement at week 52; mean change in PGA score at week 24, percentage of patients with prednisone reductions ≥ 25% from baseline to ≤7.5 mg/day during weeks 40 to 52 (in subjects whose baseline dose was > 7.5 mg/day); mean change in SF-36 PCS at week 24.

The major secondary outcome of percentage of SRI responders at week 76 failed to reach statistical significance (odds ratio and P value not submitted; odds ratio 1.31, 95% CI: 0.92 – 1.87, P = 0.1323 by logistic regression, taken from the FDA briefing package).[3]

Mean change in PGA score at week 24 was defined as a major secondary outcome. For the whole population in BLISS-52 the change in PGA score (week 24 relative to baseline) for both groups indicated disease improvement and was greater in the belimumab group (-0.54) than placebo group (-0.39; P = 0.0003 in support of belimumab). For BLISS-76 the difference between groups was very small and in favour of placebo (-0.49 placebo and -0.48 belimumab) and did not reach statistical significance (P = 0.7987). For the Target population pooled across trials belimumab delivered a greater reduction in PGA score than placebo (P = 0.028 with mean changes of -0.42 and -0.52 for placebo and belimumab, respectively). Target population data was not been provided for the change in PGA score separately for the BLISS-52 and BLISS-76.

Components of the SRI at Week 52

A further major secondary outcome was the percentage of patients at week 52 that achieved a SLEDAI score reduction of ≥ 4 points. Both trials delivered a significantly greater percentage for belimumab than for placebo (P = 0.0024 and P = 0.0062 for BLISS-52 and BLISS-76, respectively). Similarly, the Target population data delivered a significantly greater percentage for belimumab (P = 0.0004 and P = 0.0063 for BLISS-52 and BLISS-76, respectively).

Results at week 52 for the other two SRI components (i.e. the BILAG index and PGA score) were submitted (non-major secondary outcomes). The percentage of patients in the whole population that satisfied BILAG and PGA criteria in BLISS-52 was greater for belimumab than placebo (significant at P = 0.0181 and P = 0.0048 for BILAG and PGA, respectively); however, for BLISS-76 the differences between belimumab and placebo were smaller and neither component reached statistical significance in favour of belimumab (P = 0.319 and P = 0.1258 for BILAG and PGA, respectively). Similarly, the percentage of patients in the Target population which satisfied BILAG and PGA criteria in BLISS-52 was greater for belimumab than placebo (P = 0.0099 and P = 0.0063 for BILAG and PGA, respectively); however, for BLISS-76 the differences between belimumab and placebo were far more modest and did not reach conventional statistical significance (P = 0.1297 and P = 0.1312 for BILAG and PGA, respectively).

In summary, in BLISS-52 for the total population and for the high disease activity subgroup, belimumab at 10mg/kg delivered significantly more responders at 52 weeks than placebo for

SLEDAI score reduction of≥ 4 points, no worsening in PGA, and no worsening in BILAG. However, for BLISS-76 at 52 weeks total population and high disease activity subgroup, a significant response with belimumab 10mg/kg compared to placebo was only seen with the 4-point reduction in SELENA-SLEDAI component (difference between belimumab and placebo = 22%, odds ratio = 2.4 [95% CI: 1.3, 4.4; P < 0.0063]).

Reduction in steroid usage

Reduction in steroid use was specified as a major secondary outcome. In BLISS-52 at baseline 68.6% of patients were receiving ≥ 7.5 mg/day prednisone. The corresponding percentage for BLISS-76 was 44.9%. The percentage of these patients whose steroid use was reduced in weeks 40 to 52 by the pre-specified amount was greater in the belimumab arm than the placebo arm in both trials. The difference (belimumab vs. placebo) failed to reach statistical significance in either trial: 18.6% vs. 12.0% in BLISS-52 (P = 0.0526 from logistic regression including baseline covariates) and 16.7% vs. 12.7% in BLISS-76 (P = 0.5323). For the Target population pooled across trials 15.9% and 7.1% reduced steroid use in belimumab and placebo groups, respectively (P = 0.0389 from logistic regression). For the Target population in the BLISS-52 trial there was a large difference in reduced steroid use between belimumab and placebo groups (18.5% and 5.3% respectively; odds ratio = 4.11, 95% CI: 1.29, 13.2; P = 0.0171). For the Target population in the BLISS-76 trial there was no difference between groups (11.1% and 10% reduced steroid use in belimumab and placebo groups respectively; odds ratio = 0.88, 95% CI: 0.21, 3.60; P = 0.8586).

Quality of life

The mean change in SF 36 PCS scores was specified as, a major secondary outcome. At week 24 relative to baseline it showed little difference between belimumab and placebo groups in BLISS-52 (P = 0.8870), or in BLISS-76 (P = 0.6601), or in the Target population pooled across trials (P = 0.4276).

Change in SF 36 PCS scores at week 52 was also specified as a non-major secondary outcome. No significant improvement was observed for BLISS-76 or Target populations (P = 0.5134 and P = 0.1124, respectively) however in BLISS-52 there was a difference between belimumab and placebo arms (4.18 vs. 2.96) (P = 0.0247).

In BLISS-52 over the course of the study there was a statistically non-significant difference (P value not submitted) in favour of belimumab relative to placebo in the absolute change of EuroQoL 5 Dimensions (EQ-5D) score from baseline, however the results for belimumab and placebo groups in BLISS-76 were indistinguishable. For the pooled Target population the

difference between 10mg/kg and placebo groups reached statistical significance in favour of belimumab at week 24 ($P \leq 0.01$), but the difference almost completely faded by week 52.

SLEDAI flare index

Other specified non-major secondary efficacy outcomes for which results were submitted included: time to first SLE flare (assessed using the SLEDAI Flare Index which categorizes flares as "mild or moderate" or "severe" based on 6 variables (see Appendix 1); disease progression at week 52 relative to baseline assessed using the SLICC/ACR index; fatigue over the course of the study estimated using the FACIT-Fatigue scale4 which ranges from 0 to 52 (0 is the worst possible score and 52 is the best).

In BLISS-52 the time to first flare and time to first severe flare were delayed by belimumab relative to placebo (HR 0.76, 95% CI: 0.63 – 0.91, P = 0.0036; HR 0.57, 95% CI: 0.39 – 0.85, P = 0.0055, respectively). In BLISS-76 there was no difference between groups in time to first flare (P = 0.4796) but relative to placebo belimumab somewhat delayed time to first severe flare (HR 0.72, 95% CI: 0.50 – 1.05, P = 0.0867). For the Target population pooled across trials, relative to placebo, belimumab significantly delayed time to both first flare (P = 0.007) and to first severe flare (P = 0.0028).

SLICC/ACR organ damage

There was little difference between placebo and belimumab groups in terms of change in SLICC/ACR score at week 52; precise values by trial were not submitted. Data reported elsewhere[5] were: BLISS-52 score change 0.06 and 0.04 for placebo and belimumab groups respectively, P for difference 0.4222; BLISS-76 score change 0.05 and 0.03 for placebo and belimumab groups respectively, P for difference 0.3415.

Fatigue

At week 52 relative to baseline the belimumab group had greater improvement in FACIT-Fatigue score than the placebo group (4.8 belimumab and 2.1 placebo in BLISS-52, $P < 0.001$; 4.6 and 2.9 in BLISS-76, P = 0.05). For the Target population at weeks 8 and 12 the difference between groups was statistically in favour of belimumab ($P < 0.05$) however the difference between groups then diminished and at week 52 the difference no longer reached conventional statistical significance.

Modified SRI response

The results for a non-pre-specified secondary outcome, the "modified SRI" at week 52, were submitted. In the "modified SRI" serological improvements (2 points each for anti-dsDNA antibodies and for complement) were not allowed to contribute toward a ≥4 points redu ction

in SLEDAI score. In BLISS-52 belimumab delivered a greater percentage of responders than did placebo (P = 0.0038); in BLISS-76 the difference in favour of belimumab failed to reach the conventional level of statistical significance (P = 0.064).

1.2.3 Safety

The submission pooled results from three randomised controlled trials (RCTs): BLISS-52, BLISS-76 and LBSL02, providing information on 675 patients who received placebo and 1458 who were exposed to belimumab. LBSL02 lasted 52 weeks, preceded the BLISS trials, was conducted in North America (98% patients from the USA) randomised 449 patients to one of four treatments: SoC + placebo, SoC + 1mg/kg belimumab, SoC + 4mg/kg belimumab, and SoC + 10mg/kg belimumab. Although all patients had a history of auto-immunity, at recruitment 30% currently lacked anti-nuclear antibodies. This trial did not employ the SRI composite outcome.

Deaths

There were 15 deaths during the controlled phase of the three trials; 3 in the placebo group (n=675), and 12 in the belimumab groups (n=1458) with 6 each in the 10mg/kg and 1mg/kg groups respectively. One death in the 1mg/kg belimumab group followed 15 weeks after the patient stopped belimumab treatment. The causes of death were various.

Adverse events

In all treatment groups > 90% of patients experienced at least 1 adverse event (AE). The most commonly occurring AEs were headache, upper respiratory tract infection, arthralgia, nausea, urinary tract infection (UTI), diarrhoea and fatigue.

The percentage of patients experiencing at least one serious AE and at least one serious AE was very similar between placebo and belimumab groups ranging from13.5% to 18.6%, with a very slight numerical excess in the belimumab group. The most frequent serious AEs (≥ 1% in any treatment group) were pneumonia, pyrexia, UTI, cholelithiasis, and cellulitis. The percentage of patients experiencing at least one severe AE was 15.4% for the placebo group and 16% across the belimumab groups; the most common severe adverse events were not identified.

Occurrence of infusion plus hypersensitivity reactions was similar between belimumab and placebo-treated patients (17% and 14.7%, respectively).

Infections

Infections occurred slightly more often in patients treated with belimumab compared to placebo. The most frequent infections were upper respiratory tract infection (URTI), UTI, nasopharyngitis, sinusitis, and bronchitis.

Malignancy

Five solid organ malignancies were reported across the trials: a stomach carcinoid (placebo group, day 202); a breast cancer (belimumab 1mg/kg, day 102); a cervical cancer (belimumab 1mg/kg, day 439); an ovarian cancer (belimumab 1mg/kg, day 21, patient died); and a thyroid cancer (belimumab 1mg/kg, day 378). There were four non-melanoma skin cancers: two basal cell carcinomas, and two squamous cell carcinomas (1 in the placebo group, 3 in the belimumab 1mg/kg group).

1.3 Summary of the ERG's critique of clinical effectiveness evidence submitted

The submission omitted results for the 1mg/kg groups from the two pivotal trials. Therefore from information in the manufactures submission (MS) alone, the consistency of results across the whole data set could not be fully assessed and it was not possible to gauge the evidence for a dose response relationship. However, data for the 1mg/kg groups is available in the public domain (FDA documents pertaining to the Human Genome Sciences (HGS) Briefing Document to the FDA[3,5]) and the ERG have considered this information in critiquing the submitted evidence.

Even without the 1mg/kg group results the MS provided clinical evidence for a large number of outcomes reported for six separate populations (whole population from BLISS-52, whole population from BLISS-76, pooled whole populations from BLISS-52 plus BLISS-76, pooled Target populations from BLISS-52 plus BLISS-76, and after the clarification process Target population from BLISS-52 and Target population from BLISS-76. Additionally, AEs for LBSL02 were included.[6]

The most noticeable aspect of the submitted results was the relative lack of evidence for clinical effectiveness of belimumab seen in the BLISS-76 trial. Although at week 52 for the pre-specified primary outcome measure the percent responders (SRI) reached statistical significance in favour of belimumab ($P = 0.027$), at no other time point did this outcome reach significance. Furthermore, all major and non-major secondary outcome results submitted, except for a ≥ 4 point SLEDAI improvement at week 52 which is a subcomponent of the SRI response, likewise failed to reach statistical significance including: PGA change at week 24 and 52, SRI responders at week 76, reduction in use of steroids week 40 to 52, SF-36 change

at week 24, time to first flare, time to first severe flare, change in SLICC/ACR organ damage score at week 52, fatigue status (FACIT change from baseline), and quality of life (EQ-5D change).

The SLE population in BLISS-76 is more likely to resemble that in the UK than that in the BLISS-52; therefore the BLISS-76 results are probably more relevant to the decision problem than those from BLISS-52. The results favourable for belimumab submitted for the whole population pooled across trials were largely driven by BLISS-52 results. For the Target population the results from the BLISS-52 trial were again more favourable to belimumab than those from BLISS-76 and additionally BLISS-52 provided more patients to the pooled Target population than BLISS-76 (55% vs. 45%), therefore results favourable to belimumab for the pooled Target population were again more strongly driven by the contribution from the BLISS-52 Target population.

Results in the public domain[3,5]for the 1mg/kg and 10mg/kg dose regimens in the BLISS-76 trial were not supportive of a dose response relationship. For many outcomes the results were as favourable for the low dose group as for the high dose group. These outcomes included: percentage of SRI responders across trial follow up, percentage of patients with no worsening in PGA at week 52, percentage of patients with no worsening in BILAG index at week 52, mean change in PGA score from baseline at week 52, reduction in steroid use weeks 40 to 52, time to first flare, time to first severe flare, and mean change in FACIT fatigue score at week 52.

1.4 Summary of cost effectiveness submitted evidence by the manufacturer

No published relevant economic evaluations were identified in the submission. The search strategy was poor but it appears unlikely that economic studies were missed.

The submitted cost-effectiveness work focuses entirely on a new model and economic evaluation undertaken by the manufacturer. This de novo individual patient micro-simulation model examined the cost-effectiveness of belimumab plus SoC versus SoC. This employed an annual cycle over a lifetime horizon and conformed to the NICE reference case.

In brief, the model was constructed using three main sources of data:

- The BLISS trials
- The John Hopkins University cohort
- Additional data drawn from the wider literature

The trial data determined distribution of patient characteristics at baseline. Regression analysis was also used to model patients' SLEDAI score at week 52. Those in the belimumab arm who responded at week 24 were modelled as remaining on belimumab and maintaining the modelled SLEDAI score at week 52. Those in the belimumab arm who did not respond at week 24 were assumed to stop treatment and were modelled as having the average SoC SLEDAI score at week 52. Discontinuation rates from week 24 were also drawn from the trial data.

Regression analysis from the trial data was also used to derive HRQoL and cost functions related to a patient's SLEDAI score.

Given these inputs, the bulk of the remainder of the model was derived from the Johns Hopkins University (JHU) cohort data. The survival function and the risks of developing each of the 12 organ involvements within the SLICC index were modelled on a range of covariates, these including the adjusted mean SLEDAI score to date and the average cumulative prednisolone dose. Steroid use was not drawn from trial data but was rather modelled using a function estimated from the JHU cohort relating steroid use to a patient's SLEDAI score. The evolution of the SLEDAI score subsequent to week 52 was also estimated from the JHU cohort data, with the manufacturer adjusting the constant of the functional form to better fit the belimumab phase II trial data.

The survival function estimated from the JHU cohort was amended by SMRs drawn from a paper within the literature. Additional data drawn from the literature was used to inform the HRQoL and cost impacts arising from involvement of the individual 12 organ involvements within the SLICC index.

Base case deterministic results were submitted by the manufacturer for three patient populations:

- The All BLISS patient population;
- The patient population within the BLISS trials that relates to the anticipated license of Anti-dsDNA+ve and low (C3 or C4);
- The Target population which restricts the patient population to the licensed patient population with an SS score at baseline of at least 10.

For the All BLISS population the central survival estimate was an additional 1.50 years survival from use of belimumab. The discounted net gains and costs were 0.43 QALYs at a

net cost of £35,584 to yield a cost effectiveness estimate of £82,909 per QALY. With the PAS the net cost fell to £[redacted] to yield a cost effectiveness estimate of £[redacted] per QALY.

Only the base case deterministic results were presented for the anticipated license population. The central survival estimate was an additional 2.13 years survival from use of belimumab. The discounted net gains and costs were 0.61 QALYs at a net cost of £40,303 to yield a cost effectiveness estimate of £66,170 per QALY. With the PAS the net cost fell to £[redacted] to yield a cost effectiveness estimate of £[redacted] per QALY.

For the Target population the central survival estimate was an additional 2.93 years survival from use of belimumab. The discounted net gains and costs were 0.81 QALYs at a net cost of £51,925, to yield a cost effectiveness estimate of £64,410 per QALY. With the PAS the net cost fell to £[redacted] to yield a cost effectiveness estimate of £[redacted] per QALY. The direct costs of belimumab and its administration accounted for 90% and 17% respectively of the total net costs. Lower costs from pulmonary involvement and from renal involvement provided costs offsets of around 6% and 4% respectively.

An ERG cross-check of the probabilistic modelling for the Target population resulted in a central estimate of £65,530 per QALY. Due in part to the results being reasonably linear and also the time inherent in running the model probabilistically all other results presented are from the deterministic model.

A range of sensitivity analyses were presented for the All BLISS patient population and the Target population. Restricting attention to the Target population, results were sensitive to the initial changes in the SLEDAI score that were modelled, the manufacturer adjustment to the long term SLEDAI score function, the impact of the adjusted mean SLEDAI score upon mortality and the natural history models for pulmonary and renal involvement.

1.5 Summary of the ERG's critique of cost effectiveness evidence submitted

Assuming that belimumab week 24 non-responders will experience the average SLEDAI score within the SoC arm is likely to have over-estimated the average impact upon SS scores within the belimumab arm. The SLEDAI score drives the analysis and any error in its calculation will have a major impact on results. There may also be errors in the calculation of the adjusted mean SLEDAI score from not taking into account a patient's probable prior history, with this concern also applying to the calculation of the cumulative average steroid dose.

The maintenance through time of the absolute gain in SLEDAI score among those remaining on treatment compared to those on SoC may be optimistic, and at a minimum should have been explored in a scenario analysis. However, ERG expert opinion indicates that it may in some sense be possible to "reset" the immune system which may negate this concern.

Whether it is reasonable to extrapolate the 8% annual discontinuation rate for the Target population beyond the trial period is also unclear given the lack of detail around the figures underlying this rate. A high discontinuation rate from week 24 improves the cost effectiveness of belimumab.

Adjusting the JHU cohort survival model by SMRs from the literature may not be justified. The SMRs applied may also not be representative of the overall literature. This may have tended to exaggerate the impacts of the covariates within the JHU cohort survival model.

Costs as a function of the SLEDAI score may have been exaggerated by analysing the data on a six monthly basis rather than the annual basis on which the model is constructed. There are also some concerns that the separate estimation of costs for each organ involvement may have tended to double count the cost impacts of SLE.

There appear to be some discrepancies in the reported model outputs between the average durations of organ involvement, the annual costs of these and the discounted total costs of these organ involvements. There are as a consequence concerns around the calculation of the cost offsets from reduced organ involvement arising from belimumab.

With the exception of the last point the effects of which are currently ambiguous, the above suggest that the model may have tended to overestimate the impact of belimumab on the SLEDAI score and to have overestimated the likely impact of the SLEDAI score on the model outputs. There are few immediately obvious biases pulling in the opposite direction.

1.6 ERG commentary on the robustness of evidence submitted by the manufacturer

The submitted evidence concerned the clinical and cost effectiveness of 10mg/kg belimumab, used as an add-on to standard of care, compared to standard of care alone. Several SLE populations were considered. Evidence for clinical effectiveness came from two placebo-controlled phase III trials: BLISS-52[7] was conducted at 90 centres located in Pacific-Asia, South America and Eastern Europe (11 centres); BLISS-76 was conducted in 136 centres in

North America and Europe. As such the results from BLISS-76 are more likely to be generalisable to the UK.

Total populations in the BLISS trials (auto-antibody positive patients with active SLE disease sufficient to score ≥ 6 points on the SLEDAI scale) conformed to that in the NICE scope ,but the effectiveness of belimumab in the two trials was disproportionately greater in BLISS-52 than BLISS-76 and the evidence for effectiveness from BLISS-76 was not convincing. The manufacturer pooled data from the two BLISS trials, but the pooled analyses that favoured belimumab were almost exclusively driven by the effectiveness results from BLISS-52 and are arguably less applicable to the UK than the BLISS-76 results alone. The results for PGA of disease activity were noticeably disparate between trials. The ERG considered that inadequate allocation concealment of outcome assessors (physicians) in BLISS-52 might explain this discrepancy and may be a cause for concern since PGA is a component of the composite primary outcome.

The manufacturer submitted evidence for a high disease activity sub-population from the BLISS trials; this was called the "Target population". The Target population was not a pre-specified subgroup in the trial protocols; it was identified using post hoc analyses to seek out a more strongly responding subgroup of patients. The Target population was defined according to baseline disease activity score ≥ (10 SLEDAI points), l evel of antibodies to dsDNA, and low level of complement. Each of these three criteria defined a pre-specified subgroup from the BLISS trials, but the combination of all three was not pre-specified. The Target population represents a subpopulation of the proposed licensed population which in turn is a subpopulation of the total BLISS trial population. For the Target population only outcome results pooled across the trials were submitted and it was impossible to check for consistency of results between trials. This was a cause for concern because of the lack of convincing evidence of effectiveness for the whole population in BLISS-76 (see above).

Both BLISS trials had three randomised groups: placebo, 1mg/kg belimumab and 10mg/kg belimumab. The submission did not include results for the 1mg/kg arms of the trials; however these results are available in the public domain (FDA documents[3,5]). Results for the 1mg/kg dose regimen are relevant for checking consistency between trial results and in determining whether a dose response relationship exists. In this respect the most noticeable result was that in BLISS-76 there was essentially no evidence of any dose response relationship across the time span of the trial and no difference between the proportions of SRI responders to 1mg/kg and to 10mg/kg. For the SRI subcomponents PGA and BILAG in BLISS-76, the 1mg/kg regimen was more effective than 10mg/mg; similarly the 1mg/kg dose regimen appeared as effective as 10mg/kg in suppressing flares.

1.6.1 Strengths

The main strengths of the clinical effectiveness evidence were:

- The significant result ($P < 0.05$) for the pre-specified primary end point (52 week SRI) in both the two phase III RCTs;
- The fact that this primary outcome had been developed in consultation with a licensing authority (the FDA) and guarded against the possibility that improvement in some particular SLE manifestation or manifestations might mask deterioration in overall disease activity or involvement of new organ damage.

The main strengths of the cost effectiveness submission were:

- Provision of a well constructed model which conforms to the NICE reference case;
- An impressive attempt at modelling the longer term effects of SLE using extensive modelling of the JHU SLE cohort;
- The presentation of a simple and transparent PAS that allows easy implementation within the economic model.

1.6.2 Weaknesses and areas of uncertainty

There were a number of weaknesses and uncertainties; these include:

- The lack of convincing evidence from BLISS-76 that belimumab is superior to placebo in the total population. Only at week 52 did the proportion of SRI responders reach statistical significance in favour of belimumab. At other monitoring times significance was not reached and results for 1mg/kg (available from FDA reports[3,5]) and 10mg/kg belimumab are indistinguishable. Furthermore, of the five pre-specified major secondary end points in only one was belimumab favoured at a level that might not reasonably be accounted for by chance; MC this outcome, a reduction in the SLEDAI score of $\geq$ 4 points at week 52, represents the smallest disease activity improvement that can be considered clinically significant. None of the other major secondary outcomes in the BLISS-76 trial favoured belimumab at a level that strongly excluded the possibility of a chance result, including: PGA change at week 24 and 52, SRI responders at week 76, reduction in use of steroids week 40 to 52, SF-36 change at week 24. Furthermore, none of the other submitted secondary outcomes strongly

excluded chance from accounting for results in favour of belimumab, including: time to first flare, time to first severe flare, change in SLICC/ACR organ damage score at week 52, fatigue status (FACIT change from baseline), and quality of life (EQ-5D change). For some of these outcomes there was little distinction in effectiveness between 1mg/g and 10mg/kg belimumab dose regimens.

- There were considerable (and significant) geographical and racial differences between the BLISS trials (which may indicate potential differences in practice and in standard of care). The BLISS-76 population is more closely comparable to that of the UK than the BLISS-52 as also are the likely underlying care patterns. The submission pooled the two BLISS trials (52 and 76). The pooled results favourable for belimumab were almost exclusively driven by BLISS-52. The relevance of the pooled results for England and Wales is therefore uncertain. Similarly subgroup analysis (Manufacturer's clarification document) of the primary outcome according to geographical region (USA/Canada, Western Europe, Eastern Europe, America-excluding USA/Canada, Asia) indicated that response was strongest in America-excluding-USA/Canada and weakest in USA/Canada and Western Europe.

- The submission excluded results for the 1mg/kg arms of the two BLISS trials. The trial results available from the FDA indicated a lack of convincing evidence for an expected dose response relationship, with no consistent additional benefit from 10mg/kg compared to a 1mg/kg dose.

- The original submission only presented pooled effectiveness results for the Target population. The NICE submission template specifically requests separate results by trial when more than one trial is available. In the light of the relative lack of effectiveness displayed in BLISS-76 for the whole trial population, the lack of trial specific data for the Target population weakened the submission's case. Trial specific results for some outcomes for the Target population were made available during the clarification process. No data is available for the effectiveness of the 1mg/kg dose in the Target population.

- The ERG found outcome data common to BLISS and rituximab trials which was not explicit in the MS, so that the existence of data for an indirect comparison of interventions was not acknowledged. NICE request a rationale for not conducting meta-analysis when more than two RCTs are available. No rationale was provided in the MS.

- The BLISS populations exhibited a narrow range of SLE manifestations (mainly restricted to musculoskeletal and cutaneous problems). The BLISS trials were underpowered to estimate the effectiveness of belimumab with respect to manifestations in other domains. Also there was a lack of controlled evidence on the effectiveness of belimumab relative to SoC in the longer term beyond 76 weeks. Yet, in the economic analysis, effectiveness data from the pooled BLISS populations have been used in modelling belimumab's effect on a wide range of organ systems in SLE over a life-time horizon.

- The economic model generated a survival benefit for belimumab over SoC: an additional 2.93 undiscounted year's survival from belimumab within the Target population. There is no direct clinical evidence to support this. Actually, during the randomised phases of the belimumab trials, there were a greater number of deaths associated with belimumab than with placebo.

- The economic model generates better survival for patients with high disease activity than for those with low disease activity. This counterintuitive result appears to reflect the larger proportion of young patients in the Target population from the pooled BLISS trials. As such this will merely reflect the exigencies of trial recruitment and it is very uncertain whether this population is representative of high disease activity patients in England and Wales.

- In the economic model there may be an element of double counting in estimating the costs of complications, these costs being a function of the SLEDAI score with further costs being added for the individual components of the SLICC index.

- The economic model data from the JHU cohort to estimate a number of functions within the model: the long term evolution of the SLEDAI score, steroid use, mortality risks and the risk of developing organ involvement. The level of disease activity in this cohort is very much lower than that of the Target population and as a consequence the manufacturer made an informal adjustment to the SLEDAI score evolution function. This adjustment improved the estimate of the cost effectiveness of belimumab. Some informal justification for this adjustment has been provided within manufacturer responses to ERG clarification questions, but uncertainty remains because the reliability and validity of the adopted adjustment was not fully explored or robustly defended.

- The submission did not provide data about maintenance of response at the patient level. The SRI outcome was reported as a group response; the graph line showing percentage of responders across the duration of the trials rose and fell at various follow up times, thus an individual non-responder could later improve sufficiently to be classified as a responder.

- The economic model overestimated health benefit in the belimumab arm between weeks 24 and 52. This was because the estimate (i.e. the decrease in SLEDAI score) for non-responders in the belimumab arm (33% of patients in the belimumab arm) was calculated from observed changes in the whole SoC arm which included responders as well non-responders. In the pooled Target population the SoC arm consisted of 52% responders and 48% non-responders with average SLEDAI improvement of 6.9 points for responders and 1.1 points for non-responders; thus the improvement for the SoC arm as a whole was heavily weighted by the SoC responders. A more appropriate procedure would be to base the estimate of the health benefits for non-responders in the belimumab arm on the SLEDAI change observed for non-responders in the SoC arm (i.e. assuming that the two sets of non-responders experience similar disease trajectories as a result of their SLE). This weakness extends beyond week 52 because the manufacturer's model for the belimumab arm non-responders after week 52 continues to make an estimate of disease activity based on the whole SoC arm (made up of a mix of both responders and non-responders).

1.7 Summary of additional work undertaken by the ERG

The ERG undertook substantial additional work in the following areas:

1. Supplementing the MS with data in the public domain e.g. as available from the FDA.

2. Extensive clarifications required to understand the anticipated effectiveness in the different relevant sub-populations: Target; high disease activity and license.

3. Re-running the search strategies.

4. Running a probabilistic sensitivity analysis (PSA) to cross check the model.

1.8 Key issues

The proposed licensed population and the high disease activity "Target" population, the focus of the clinical effectiveness and economic submissions, were subgroups identified from post hoc analyses aimed at identifying patients with the greatest response to belimumab in the

pooled phase III trial populations. Although important as subgroup analyses, the results should be viewed with some caution and not assumed to represent the outcome of an independent randomised investigation of pre-defined "Target" populations.

The MS and clarification document presented results for multiple populations (whole population from two RCTs, pooled whole populations, pooled Target populations and Target population separately by trial). The results from the BLISS-76 trial, which were less supportive of belimumab than those from BLISS-52, are those more generalisable to the UK; the economic model employed results pooled across both trials and therefore may somewhat overestimate the cost effectiveness of belimumab for the UK population.

The submission did not present results for the 1mg/kg groups in the two pivotal trials. Data in the public domain,[3,5] although not formally tested statistically, indicated that for several outcomes, including the primary outcome in the BLISS-76 trial, there was little difference between the effect of belimumab in 1mg/kg and 10mg/kg groups.

The manufacturer's estimate of the number of Target population patients in the UK was based on the proportion of such patients at baseline in the BLISS trials; this will probably be an underestimate because SLE is a relapsing and remitting disease and the number of patients likely to reach Target population status at some stage in their disease will accumulate through time.

The manufacturer's economic model relied heavily on time to event analyses of the JHU SLE cohort. Based on SELENA SLEDAI scores there was a gross mismatch between JHU cohort patients and the populations modelled, the former had far less severe disease, especially in comparison to the Target population. To allow for this mismatch a major adjustment was required in modelling; the manufacturer's justification for the type of adjustment adopted was that a similar procedure had been explored in cardiovascular studies for the prevention of cardiovascular disease (CVD).[8] The robustness of the manufacturer's approach in this respect is difficult to gauge.

Participants in the 52 and 76 week pivotal RCTs experienced a relatively narrow range of SLE manifestations, predominantly in musculoskeletal, cutaneous and serological domains. The economic analysis used trial changes in SELENA SLEDAI scores for these patients in order to model long term accumulation of organ damage in many other systems. The reliability of this procedure is again difficult to gauge and was referenced in the MS with a single analysis published in 1999.[9]

2 BACKGROUND

2.1 Critique of manufacture's description of underlying health problem

The manufacturer provides an adequate description of the treatment pathways for patients with SLE. These are presented MS section 2 (Page 39).

The ERG is of the opinion that the manufacturer's summary of the disease context and available treatments for patients with SLE is reasonable.

2.2 Critique of manufacturer's overview of current service provision

The MS executive summary and MS section 2 adequately describe the aims and modes of treatment (Pages 17 to 18). The key points, taken from the MS, are as follows:

Treatments aim to: match treatment to an accurate diagnosis of the extent of organ involvement; maintain an appropriate level of therapy to control or halt the inflammatory disease activity while minimising side-effects and risk of infection; prevent further organ damage; maintain a patient's daily function and quality of life.

Currently a range of treatments (including NSAIDs, corticosteroids, immunosuppressants and antimalarials) are variously used either alone or in different combinations, constituting standard of care (SoC). The MS documents that current SoC may be associated with undesirable effects e.g. from chronic use of steroids or side effects associated with immunosuppressants. The MS points out that many treatments are not licensed for use in SLE and that a significant number of patients with advanced SLE do not respond to current treatments even at high doses". Patients with more severe, highly active SLE are usually managed in tertiary centres and may routinely receive rituximab.

Currently there is no accepted SLE treatment algorithm and no relevant NICE guideline exists. Agreeing on best practice poses a significant challenge owing to the heterogeneous nature of SLE.

The manufacturer has estimated the number of patients presenting with SLE in England and Wales who would be eligible for treatment with belimumab (see MS Table 2.2). Taking 13,198 as the number of patients with active SLE in England and Wales, the MS calculated that 92.5% of these are adults. The estimated number eligible for belimumab according to the "proposed license indication" was then based on the proportion of patients (52%) in the

pivotal phase III trials who fitted the criteria defining the "proposed license population" giving a total of 6,348 (i.e. 0.925 * 0.52 * 13,198). This proposed license population exhibited a higher level of serological disease than the total Phase III populations. However the MS further submitted that NICE should actually consider belimumab for a subgroup of the "proposed licence population". This population was a narrower population of high disease activity patients termed the "Target population", representing 34% of the patients in the Phase III trials, giving an estimated number of 4,151 patients in England and Wales (i.e. 0.925 * 0.34 * 13,198).

The manufacturer's calculations should be viewed with some caution because the Phase III trials upon which they are based were international studies in which UK patients were a very small minority, and because the actual proportions of "proposed license population" and "Target population" patients in these trials will reflect the vagaries of trial recruitment rather than the distribution of these patients in the countries from which they were selected.

The manufacturer's estimate of the cost to the NHS of treating the "Target population" in year one assumed 50% usage of belimumab (i.e. 2,075 patients) and came to [redacted] under the manufacturer's proposed patient access scheme (PAS). The manufacturer's estimate of the number of eligible patients rose by 346 in year two, and then by 137 for each of the next three years to year five in which the estimated cost to NHS was [redacted].

3 CRITIQUE OF MANUFACTURER'S DEFINITION OF DECISION PROBLEM

Table 1 shows the MS decision problem with rationale for deviations from the NICE scope.

Table 1: Manufacturer's indicated scope (from MS Table 4.1)

	Final scope issued by NICE	Decision problem addressed in the submission	Manufacturer's Rationale if different from the scope
Population	Adults with active autoantibody-positive systemic lupus erythematosus	**Phase 3 Trial Population** Adults with active autoantibody-positive systemic lupus erythematosus. **High Disease Activity Subgroup** Adults with active autoantibody-positive systemic lupus erythematosus with evidence for serological disease activity (low complement, positive anti-dsDNA) and SELENA-SLEDAI ≥ 10.	Mindful of NHS resources, the proposed population of interest to this decision problem is a subgroup of the Phase 3 trial population which applies the additional criteria of evidence for serological disease activity (low complement, positive anti-dsDNA) and SELENA-SLEDAI disease activity score of ≥ 10 This subgroup experienced an additional treatment effect to belimumab over and above the Phase 3 trial population and is aimed at identifying SLE patients at the greatest risk of experiencing long-term organ damage.
Intervention	Belimumab as an add on to standard therapy	Belimumab 10mg/kg administered as an intravenous infusion over a one hour period on days 0, 14 and 28, and at 4 week intervals thereafter in addition to standard therapy.	

	Final scope issued by NICE	Decision problem addressed in the submission	Manufacturer's Rationale if different from the scope
Comparator(s)	• Standard therapy alone For people in whom it is considered appropriate: • Rituximab plus standard therapy • Cyclophosphamide plus standard therapy	• Standard therapy which comprises (alone or in combination): antimalarials, NSAIDs, corticosteroids, or other immunosuppressants (azathioprine, methotrexate, and mycophenolate mofetil) • Rituximab plus standard therapy for the more severe SLE sub-population	Despite failing to meet primary or secondary outcomes in a Phase 2/3 SLE trial, rituximab, is used in the more severe patient population in addition to standard therapy. Therefore, rituximab plus standard therapy is a relevant comparator. The patient population and outcomes measured are not comparable to those in the belimumab trials. Therefore, conducting indirect comparisons of efficacy are problematic and have not been incorporated into the cost-effectiveness model. However, the benefits of belimumab compared with rituximab will be discussed in the written submission. Cyclophosphamide, whilst used in the more severe patient population, is largely reserved for the treatment of lupus nephritis. This is not the proposed Target population for belimumab, therefore, cyclophosphamide plus standard therapy is not a relevant comparator. In addition, adverse effects associated with long-term exposure to cyclophosphamide including bladder cancer, bone marrow suppression, haematologic malignancies, infections, myelodysplasia, and infertility limit the appropriateness of cyclophosphamide given that a high proportion of patients are women of childbearing age.

	Final scope issued by NICE	Decision problem addressed in the submission	Manufacturer's Rationale if different from the scope
Outcomes	The outcome measures to be considered include: • disease activity • incidence and severity of flares • mortality • health-related quality of life, including fatigue • adverse effects of treatment	The outcome measures included in the cost-effectiveness model are: • Disease activity • Incidence and severity of flares • Mortality • Health-related quality of life • Disease progression in terms of long-term organ damage – As discussed at the scoping workshop, although not collected in the clinical trials, long-term organ damage will be considered in the assessment of cost-effectiveness based on modelled data from the Johns Hopkins Lupus Cohort Additional endpoints discussed in the written submission and not included in the health economic model are: • Fatigue - In the Phase 3 trials this was measured using the FACIT-Fatigue instrument and was reported as the mean change in scale score at Weeks 12, 24, 52 and 76 (BLISS-76 only) • Adverse events of treatment	Adverse effects of treatment have not been included in the base case economic model as significant differences between treatments were not noted from the two pivotal Phase 3 trials. The side effect profile of belimumab will be discussed in the clinical section of the submission.

	Final scope issued by NICE	Decision problem addressed in the submission	Manufacturer's Rationale if different from the scope
Economic analysis	The reference case stipulates that the cost effectiveness of treatments should be expressed in terms of incremental cost per quality-adjusted life year. The reference case stipulates that the time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared. Costs will be considered from an NHS and Personal Social Services perspective.	• Cost effectiveness will be expressed in terms of incremental cost per quality-adjusted life year • The time horizon for the model will be lifetime • Costs will be considered from an NHS and Personal Social Services perspective	Not applicable.
Subgroups to be considered	None outlined in scope.	See population section above.	See population section above.
Special considerations, including issues related to equity or equality	None outlined in scope.	It will be important to acknowledge the innovative nature of belimumab in the treatment of SLE. There is a limitation with the current cost per QALY methodology not able to capture all the benefits of belimumab (i.e. avoidance of corticosteroids, impact of fatigue and loss of productivity). SLE has a significantly greater impact on certain ethnic groups and is most prevalent in woman of childbearing age.	

3.1 Population

The manufacturer's scope specified two populations: the Phase III trial population (adults with active autoantibody-positive SLE), and a High Disease Activity Subgroup (HDAS).

The submitted clinical effectiveness evidence came from two multicentre international Phase III RCTs (BLISS-52 and BLISS-76). The geographical location of study centres differed considerably between trials. In BLISS-52 there were 90 centres: in 13 countries in Latin America there were 38 centres (Argentina, Brazil, Chile, Colombia and Peru), in Asia-Pacific there were 41 centres (Australia, Hong Kong, India, Korea, Philippines and Taiwan) and in Eastern Europe there were eleven centres (Romania and Russia). In BLISS-76 there were 136 centres in 19 countries in North America (Canada, Costa Rica, Mexico, Puerto Rico and USA) and Europe (Austria, Belgium, Czech Republic, France, Germany, Israel, Italy, The Netherlands, Poland, Romania, Slovakia, Spain, Sweden and UK); North America (65 centres) and Europe (62 centres) contributed 93% of the centres in BLISS-76. These geographical differences were reflected in racial differences between the populations in the two trials. Although both trials included adults with auto-antibody positive active SLE it appears clear that the population in BLISS-76 is more likely to be similar to that in England and Wales than that from BLISS-52. It is reasonable to assume that the results from BLISS-76 will be more generalisable to the UK. This would be of little consequence if the clinical results were consistent between trials; however this was not so for some outcomes and in general very little clinical benefit was observed in BLISS-76 compared to some benefits in BLISS-52.

The manufacturer's scope also specified a HDAS termed the "Target" population and described as the focus of the submission. The identification of the Target population, and the evidence for clinical effectiveness of belimumab in the Target population, arose from post hoc analyses of the two BLISS trials. The rationale for this deviation from NICE scope was largely on economic grounds in that cost effectiveness was more favourable. Because the BLISS-76 trial subpopulation is more likely to match high disease activity patients in the UK than the BLISS-52 subpopulation, it is arguable that the BLISS-76 Target population is the most appropriate.

The Target population was defined as: *"Adults with active autoantibody-positive systemic lupus erythematosus with evidence for serological disease activity (low complement, positive anti-dsDNA) and SELENA-SLEDAI ≥ 10"* [MS Page 53]. There are undoubtedly patients in the UK who correspond to the Target population; however, according to expert clinical opinion, the SELENA SLEDAI is not commonly used to define high disease activity and it

may be difficult to estimate the number of patients in the UK who fit this definition. The manufacturer's estimate of 4,150 patients across England and Wales is presented on Page 310 of the submission.

The population proposed in the license application is a high disease activity subgroup, termed "population A" in the submission, and defined in the Summary of Product Characteristics (SPC) in Box 2.

Box 2: Proposed license population

"...adult patients with active, autoantibody-positive systemic lupus erythematosus (SLE) with a high degree of disease activity (e.g. positive for anti dsDNA, low complement) despite standard therapy."

The submission presented very little evidence about the effectiveness of belimumab in the proposed licensed population (only one Figure was given (MS Figure 5.3; Page 96). The submission estimated that there are 6,348 "*population A*" patients in England and Wales. The Target population represents a subpopulation (~64.5%) of population A patients.

3.2 Intervention

The intervention described in the submission matches that in the NICE final scope

Belimumab is a human IgG1λ monoclonal antibody, expressed in a murine cell line that binds to soluble human B-lymphocyte stimulator (BLyS) and inhibits its biological activity. BLyS influences differentiation, survival and activation of B lymphocytes. In the proposed license Belimumab is delivered at 10mg/Kg by a one hour IV infusion. It is an add-on therapy to standard of care that commonly consists of a range of treatments (NSAIDs, corticosteroids, immunosuppressants and antimalarials) used alone or in various combinations.

Belimumab awaits marketing approval in Europe. Application was filed with the European Medicines Agency (EMA) on 4th June 2010 and is now under review via the Centralised procedure. CHMP opinion was expected in May 2011 followed by a Commission decision on European marketing authorisation in July 2011. [redacted]

In March 2011 the USA FDA approved belimumab for reducing disease activity in adult patients with autoantibody positive SLE. This is a wider population than that encompassed in the European license application according to the SPC document submitted by the manufacturer.

3.3 Comparators

Three comparators were identified in the NICE final scope: standard of care (SoC), rituximab, and cyclophosphamide. The clinical effectiveness and economic sections of the submission did not quantitatively consider rituximab or cyclophosphamide as comparators, only SoC was formally assessed.

The MS justifies the exclusion of rituximab as a comparator on the following grounds:

- There has been no head to head trial of rituximab versus belimumab;
- Outcome measures used in rituximab and belimumab trials have differed to the extent that there is little possibility of undertaking meaningful indirect comparison meta-analysis;
- Rituximab has not been shown to be effective versus SoC whereas belimumab has, therefore by implication belimumab is unlikely to be less effective than rituximab;

[redacted]

Regarding effectiveness, although the primary end point was not reached in the Phase II/III rituximab RCT (EXPLORER)[10] the ERG's clinical expert indicated that the EXPLORER end point was more stringent than that in the BLISS trials because it registered a sustained response (once a patient was classified as a non-responder they remained so classified for the remainder of the trial), whereas the primary end point in BLISS was a group response in which a non-responder could later become classified as a responder for the primary end point at week 52.

A literature search undertaken by ERG revealed published information on SLEDAI and SF 36 changes in the EXPLORER trial which might have been used for comparison with the BLISS trials. Furthermore, RCTs for both drugs recorded BILAG changes thus offering the potential for an indirect comparison to be undertaken[10,11]. For these reasons the ERG requested clarification regarding the manufacturer's justification for not considering rituximab as a comparator.

The manufacturer responded with further justification as shown in Box 3.

Box 3: Justification for no formal comparison of belimumab and rituximab

The main reason for this decision relates to important differences in patient selection and consequently the treatment management protocols employed in the studies, …….. The patients in the rituximab Phase 2/3 trial had significant and acute disease activity at entry to the study; 53% had at least one BILAG A score (severe disease activity) and a further 28% had at least 3 BILAG B scores (please note that although a BILAG B score represents moderate disease activity, the presence of 3 BILAG B scores in some organs indicates more severe disease activity). Initially, patients were receiving very high daily doses of prednisone (mean 45.9 mg ±16.4mg) to treat the significant level of disease activity and this dose was to be tapered where possible during the trial. In addition, all patients were receiving one immunosuppressant at study entry. In contrast, the patients in the BLISS studies were a broader population and not all patients were experiencing major disease flares at study entry requiring the very high doses of steroids seen in the rituximab trial. Even in the high disease activity subgroup (Target population), only 19.3% had at least one BILAG A score at baseline, the average prednisone or equivalent dose was 12.3 mg ± 9.6mg and 53% were on an immunosuppressant. In particular, we believe that the differences in the use of steroids and immunosuppressants to manage disease activity between the rituximab and BLISS trials and consequently the differences in the type of response observed in the placebo arms render the studies incomparable.

Justification for excluding cyclophosphamide as a comparator was stated as shown in Box 4.

Box 4: Justification for no formal comparison of belimumab and cyclophosphamide

Cyclophosphamide, whilst used in the more severe patient population, is largely reserved for the treatment of lupus nephritis. This is not the proposed Target population for belimumab".

The submitted SPC is shown in Box 5.

Box 5: From the submitted SPC document

There are no or insufficient data available on the effects of Benlysta in patients with severe active lupus nephritis or severe active central nervous system lupus. Therefore, Benlysta cannot be recommended to treat these conditions.

3.4 Outcomes

The NICE and manufacturer's scopes state that the outcome measures to be considered include: disease activity, incidence and severity of flares, mortality, HRQoL including fatigue, and adverse effects of treatment. All these are reported in the MS.

The primary outcome in the BLISS trials was the proportion of responders at week 52. To estimate the proportion of responders a novel composite outcome measure, the SLE Responder Index (SRI) was introduced. The SRI was developed in conjunction with the FDA to be used in the BLISS trials. The SRI outcome aims to detect improvement in disease activity in terms of resolution of an SLE manifestation or manifestations (estimated using the SELENA SLEDAI instrument) while guarding against the possibility that this improvement might mask detrimental involvement of new organ systems (estimated using the BILAG) index) or an overall deterioration in well being (estimated using a PGA).

These three components, SELENA SLEDAI and BILAG and PGA had to be satisfied according to pre-specified requirements before a patient could be classified as a responder. These requirements were as follows:

- A ≥ 4 point reduction in the SELENA-SLEDAI score compared to baseline;
- No worsening (an increase of no more than 0.3 points) in PGA score compared to baseline;
- No new BILAG A organ domain scores or no 2 new BILAG B organ domain scores at time of assessment compared to baseline.

The SELENA SLEDAI instrument detects the presence of a manifestation of SLE disease. It encompasses 24 individual SLE manifestations, each with a weighted score from 1 to 8 points. Assessment relates to the preceding 10 days. Each manifestation must be related to lupus. A summed score of ≥ 6 across manifestations is considered active disease. A decrease of 4 points relative to previous assessment is thought to equate to a clinically meaningful improvement. For most manifestations there is no intermediate score, the item is registered as presence or absent so that a SELENA SLEDAI item generally can only improve by its resolution. The tool is therefore a measure of improvement and is not designed to assess worsening of a manifestation once present. The SRI uses the BILAG and PGA as measures of worsening.

The BILAG CLASSIC instrument includes 86 items grouped into 8 organ systems, general (5 items), mucocutaneous (18 items), neurological (15 items), musculoskeletal (9 items), cardiorespiratory (12 items), vasculitis (8 items), renal (11 items), and hematological (8 items)[2]. A score is calculated for each system depending on the SLE clinical manifestations (or signs and symptoms) present and whether they are new, worse, the same, improving, or not present in the last 4 weeks compared with the previous 4 weeks. A BILAG A score is given for a disease manifestation considered sufficiently severe to normally require high-dose steroids (prednisolone > 20 mg/day or equivalent) and/or immunosuppressive / cytotoxic agents under normal circumstances. A more moderate manifestation, which it would be considered appropriate to treat with lower dose steroids, antimalarial drugs or NSAIDs, constitutes a BILAG B score. A mild symptomatic manifestation that would require only symptomatic therapy (e.g. analgesics and NSAIDs) constitutes a BILAG C score. If there are no current symptoms, but the system has previously been involved, then a BILAG D score is recorded, while if the system has never been involved, a BILAG E score is assigned.

3.5 Economic analysis

The manufacturer's economic analysis is in line with that stipulated in the NICE scope. The MS presented its economic assessment in terms of incremental cost per QALY and has modelled outcomes using a lifetime horizon. Costs are considered from an NHS and PSS perspective.

3.6 Other relevant factors

Special considerations and issues raised in the manufacturer's scope include: 1) the innovative nature of belimumab for SLE; 2) the inability of the utility method to capture the QoL of SLE patients sufficiently sensitively; and 3) the impact of SLE on particular ethnic groups and on women of childbearing age. The proposed SPC specifies that belimumab should not be administered to pregnant women or to those planning pregnancy and therefore the special consideration relating to women of childbearing age appears to be of marginal relevance.

There were no issues identified in the NICE scope.

4 CLINICAL EFFECTIVENESS

4.1 Critique of manufacturer's systematic review

The objective of the manufacturer's systematic review was stated in Box 6 (MS Page 60).

Box 6: Objective of systematic review

A systematic review of the published literature was conducted to identify all relevant published randomised controlled trials (RCTs) for belimumab and relevant comparators in SLE.

4.1.1 Description of manufactures search strategy

Two clinical literature searches are reported in the MS; one to identify RCTs and one to identify observational studies (MS Appendix 2). The purpose of the latter search was not explicitly stated. The search strategies were of good quality (a summary of the ERG's assessment is in Appendix 2). The ERG considers it unlikely that the search would have been missed relevant studies.

The searches undertaken by the manufacturer to identify all relevant RCTs were conducted on 8th December 2010. Four electronic databases were searched (Embase, Medline, Medline In-Process, The Cochrane Library). The search strategy utilised an appropriate combination of free-text and thesaurus terms to identify the patient group (systemic lupus erythematosus), the intervention (belimumab) and the comparators. A date limit and a search filter were applied to the Embase and Medline searches to limit them to studies published after 1970 and to a particular type of evidence (RCTs), which was appropriate. The search filter used in Medline closely resembles the SIGN RCT filter[12], but misses several lines relating to publication type indicating that an old version may have been used. No language restrictions appear to have been applied. In addition to database searches, hand searching was undertaken of reference lists; the proceedings of three relevant conferences between 2006-2010 and four clinical trial registers (Clinical Trials, International Standard Randomised Controlled Trial Number (ISRCTN) Register, UK Clinical Trials Gateway, metaRegister of Controlled Trials).

The searches undertaken by the manufacturer to identify non-RCT evidence were conducted on 3rd March 2011. Four electronic databases were searched (Embase, Medline, Medline In-Process, The Cochrane Library). The search strategy utilised an appropriate combination of free-text and thesaurus terms to identify the patient group (systemic lupus erythematosus) and the intervention (belimumab). Terms to identify comparators were not included. A search filter was applied to the Embase and Medline searches to limit them to a particular type of evidence (observational studies), which was not appropriate in Medline in light of the small

number of studies retrieved before the filter was applied (66). The search filter used was the SIGN observational study filter[12] No date restrictions appear to have been applied in the search strategies themselves, although this is unclear as MS Appendix 6 states that the date span of the search is "Medline & Medline In-Process: 1950 to present day and Embase: 1980 to present day". In addition to database searches, hand searching was undertaken of reference lists; the proceedings of three relevant conferences between 2006-2010 and two clinical trial registers (Clinical Trials, UK Clinical Trials Gateway). Whilst the ERG was not able to check the search results, the search strategies were of adequate quality.

The database search alone yielded 3774 references (MS Fig 5.1). It was not possible for the ERG to attempt to reproduce the manufacturer's study selection procedure because of the large number if publications retrieved and because the description of the manufacturer's selection procedure was unclear (see below). An independent selection of studies by the ERG, effectively a separate systematic review, was not within the ERG remit or feasible within time constraints for such a large number of references.

4.1.2 Inclusion and exclusion criteria used for study selection

MS Figure 5.1 provides a flow diagram for the selection of studies. With regard to selection of studies for inclusion, the MS Page 60 states as shown in Box 7.

Box 7: MS Page 60

The inclusion and exclusion criteria were chosen to identify all relevant RCTs

Details of these criteria were not clear; they may be those in MS Table 5.1 entitled "Eligibility criteria used in search strategy" and shown in Table 2. However, since ERG could find little relationship between the criteria listed and the studies listed as included studies, the ERG considers it is possible that MS Table 5.1 (see Table 2) actually represents a summary of the search strategy, in which case a formal statement of inclusion criteria was not submitted.

Table 2: MS Table 5.1 Page 61 Eligibility criteria used in search strategy

	Clinical effectiveness
Inclusion criteria	Population - Adults (≥ 18 years) with systemic lupus erythematosus (SLE); studies were also included for SLE patients with kidney involvement - Interventions o Belimumab o Rituximab o Mycophenolate mofetil o Prednisolone and other steroids o Hydroxychloroquine and other antimalarials o Azathioprine o Cyclophosphamide o Methotrexate Outcomes - Change in SELENA-SLEDAI score (Safety of Estrogens in Lupus Erythematosus National Assessment Systemic Lupus Erythematosus Disease Activity Index) - Change in BILAG score (British Isles Lupus Assessment Group) - Change in PGA (physician global assessment scale) - Change in SLICC score (Systemic Lupus International Collaborating Clinics) - Change in number/frequency of flares - Quality of life - Reduction in steroid use - Medical resource utilisation - Fatigue (e.g. FACIT, Functional Assessment of Chronic Illness Therapy score) - Adverse events including: o Incidence and severity (grade) of all adverse events (AEs) reported o Withdrawals due to AEs o Mortality o SAEs Study design - RCT, both cross-over and parallel, blinded and open-label designs Language restrictions - Only English publications (if only the abstract was in English, this would be included)
Exclusion criteria	Population - Studies enrolling patients with only active lupus nephritis were excluded Interventions - Non-specified Outcomes - Non-specified Study design - Designs other than RCT Language restrictions - Publications in languages other than English

Thus the MS was unclear about how or if the criteria listed in Table 5.1 were actually applied to the publications retrieved from searching; for example, although a search for uncontrolled studies was undertaken one of the exclusion criteria stated in Table 5.1 is that non-RCT study

designs were to be excluded. After seeking clarification the ERG remain doubtful that the criteria from Table 5.1 were systematically applied because many studies were excluded on the basis that they lacked a requirement for patients to have active autoantibody-positive SLE or because patients were receiving azathioprine, yet active autoantibody disease is not a specified inclusion criterion in Table 5.1 and azathioprine is listed as an included intervention rather than an excluded one

In short the MS and the manufacturer's response to request for clarification fail to indicate clearly the criteria used for study inclusion and exclusion.

4.1.3 Studies included

The MS systematic review provided confusing information regarding which studies were included and which were excluded. MS Figure 5.1 and Page 62 of the submission state that 43 publications were included. These are presented in Box 8.

Box 8: Statement of the number of publications included

The number of included publications was 43 (36 full publications plus seven abstracts), including eight publications (of four trials) of belimumab and 35 publications of other interventions.

The MS provided details of only 11 publications, rather than 43, (submission Tables 5.2, 5.3 and 5.4). The ERG sought clarification and a full list of the 43 identified publications was supplied together with reasons for exclusion of excluded studies. This list is reproduced in Appendix 3. The clarification implies that of the 43 publications identified nine were classed as "included". Eight of these were publications on four industry sponsored belimumab studies (RCTs: LBSL02, BLISS-52, BLISS-76; and Phase I study LBSL01), while the ninth described an RCT of rituximab conducted in adults with moderate-to-severe active SLE[10] which was not listed in the MS as an included study.

The clarification list of 43 publications was unclear on the status of the two hydroxychloroquine publications shown in MS Table 5.3 and described therein as "linked publications of competitor drugs that were also included in the systematic review." In the clarification list as "Reason for exclusion" the entry for both studies reads "Withdrawal study in patients with stable SLE"; this may represent a reason for exclusion. The ERG searched the MS text for any further reference to these two hydroxychloroquine publications but could find none.

A publication describing the industry phase I study LBSL01 was listed as "included" but this study was not discussed in the MS. The manufacturer's stated reason for this is reproduced in Box 9.

Box 9: Reasons for not including study LBS01

As this was a small (n=70) exploratory study of limited duration, designed primarily to demonstrate safety and tolerability in humans, it does not reflect the proposed clinical use of belimumab and therefore will be excluded from further discussion.

4.1.3.1 Important included studies

No RCTs were found that compared belimumab with an alternative active intervention. The most important belimumab studies identified were three industry-sponsored RCTs conducted in adults comparing belimumab plus standard care with placebo plus standard care (trials: LBSL02[13], BLISS-52[7], and BLISS-76) together with an uncontrolled extension (LBSL99) of LBSL02. One rituximab RCT (EXPLORER trial[10]) was included in narrative discussion of potential comparators. Brief details of these studies are shown in Table 3.

Table 3: Important studies included in manufacturer's submission

ID Year ψ	Study type	Study duration	Patient Age, yr	Treatment Groups ¥	N (ITT)	Countries (% enrolled)
LBSL02 2006	Phase 2 Efficacy and Safety	52 wk	20 - 75	Bel 1mg/kg IV** Bel 4mg/kg IV* Bel 10mg/kg IV* Placebo**	114 111 111 113	USA (98%) Canada (2%)
BLISS-76 2009	Phase 3 Efficacy and Safety	76 wk	18 - 73	Bel 1mg/kg IV* Bel 10mg/kg IV* Placebo**	271 273 275	USA and Canada (53%) West Europe (25%) East Europe (11%) Latin America (11%)
BLISS-52 2009	Phase 3 Efficacy and Safety	52 wk	18 - 71	Bel 1mg/kg IV* Bel 10mg/kg IV* Placebo**	288 290 287	Latin America (50%) Asia (38%) East Europe and Australia (13%)
LBSL99 2006	Safety extension of L02			Bel 10mg/kg IV*	296	USA and Canada (100%)
EXPLORER	Phase 2/3 Efficacy and Safety	52 wk	16 - 75	Rit 1000mg# Placebo***	169 88	North America (100%)

ψ Year study subject enrolment ended
¥ All treatments were additional to standard care
* Bel = Belimumab 1, 4, or 10mg/kg administration by IV infusion on days 0, 14, 28, and every 28 days thereafter
** Placebo by IV infusion on days 0, 14, 28, and every 28 days thereafter
*** Placebo by IV infusion on days 0, 14, 167 and181
\# Rit = Rituximab on days 0 and 14

4.1.4 Details of any relevant studies that were not included

The ERG conducted a systematic search for randomised controlled trials of belimumab and of rituximab; no relevant studies additional to those included in the MS were identified.

4.1.5 Summary statement on MS systematic review

The manufacturer's systematic review was confused. Although the search strategy was of good quality the use of the retrieved references to identify relevant studies for inclusion was not well described. The ERG remains unclear regarding the methods used and the list of included studies both in the MS, and the response to request for clarification was ambiguous. Despite these non-systematic aspects, studies relevant to the decision problem have been identified and the studies presenting evidence on belimumab appears complete, although a rigorous check would require a separate and independent systematic review.

4.2 Submitted clinical evidence results

4.2.1 Scope and synopsis of the studies providing clinical evidence

Belimumab was administered as additional therapy to "standard of care" and was compared to placebo plus "standard of care". No formal comparison of belimumab vs any other active intervention (rituximab) was attempted.

Four belimumab studies provided clinical evidence: three RCTs: LBSL02, BLISS-52, and BLISS-76, and an uncontrolled study (LBSL99) that was an extension of LBSL02. MS Tables 5.4 and 5.5 provide full details of these studies.

- A phase II RCT (LBSL02) with four patient groups receiving infusions of placebo (n=113), or 1mg (n=114) or 4mg (n=111) or 10mg (n=111) belimumab/kg. A peer reviewed full publication of trial LBSL02 appeared in 2009.[13]

- Two phase III RCTs, BLISS-52 (n=865) and BLISS-76 (n=819), conducted simultaneously each with three randomised groups receiving placebo or 1mg/kg or 10mg/kg belimumab infusions. A peer reviewed full publication of the BLISS-52 trial appeared in 2011,[7] BLISS-76 has yet to appear as a peer reviewed full publication.

- LBSL99, a Phase II Continuation Study of the phase II RCT LBSL02.

Table: 4 summarises the main features of the four studies. Further details of study design and patient demography are discussed in the following section of this report.

Table: 4 Belimumab studies for safety and effectiveness evidence

ID Year*	Study type	Study duration	Patient Age, yr	Treatment Groups ¥ #	N (ITT)	Countries (% enrolled)
LBSL02	Phase 2 Efficacy and Safety	52 wk	20-75	Bel 1mg/kg IV Bel 4mg/kg IV Bel 10mg/kg IV Placebo	114 111 111 113	USA (98%) Canada (2%)
BLISS-76	Phase 3 Efficacy and Safety	76 wk	18 - 73	Bel 1mg/kg IV Bel 10mg/kg IV Placebo	271 273 275	US and Canada (53%) West Europe (25%) East Europe, (11%) Latin America (11%)
BLISS-52	Phase 3 Efficacy and Safety	52 wk	18 - 71	Bel 1mg/kg IV Bel 10mg/kg IV Placebo	288 290 287	Latin America (50%) Asia (38%) East Europe and Australia (13%)
LBSL99	Safety extension of LBSL02			Bel 10mg/kg IV	296	USA and Canada (100%)

* Year study subject enrolment ended
¥ All treatments were additional to standard care
\# Bel = Belimumab 1, 4, or 10mg/kg administration by IV infusion on days 0, 14, 28, and every 28 days thereafter

For the assessment of safety, the submission pooled data from all belimumab arms of the three RCTs (LBSL02, BLISS-76, and BLISS-52 providing data for1458 patients) and from all placebo arms (providing 675 patients).

Although all patients in study LSBL02 had a history of anti DNA-antibodies, approximately 30% lacked positive auto-antibody status at recruitment. Consequently the MS excluded this study from the clinical effectiveness analyses and it was only included for assessment of safety. For the assessment of clinical effectiveness the submission presented results from BLISS-52 and BLISS-76 phase III trials for the placebo and 10mg/kg belimumab arms only. The explanation for excluding results for the 1mg/kg dose regimen was stated in Box 10 (MS Page 102).

Box 10: Manufacturer's reason for not including results for the BLISS 1mg/kg groups

Whilst a 1mg/kg dose was examined in the Phase III studies, we will only present results for the 10mg/kg belimumab dose as this is the dose submitted for Marketing Authorisation.

Since results for the 1mg/kg arms of the trials can provide information about consistency of response and the existence of a dose response relationship, when considered relevant the ERG have made use of public domain data provided in FDA documents pertaining to the USA licensing application for belimumab[3,5].

The submission compared clinical effectiveness of 10mg/kg belimumab vs placebo for the following populations:

- BLISS-52 alone
- BLISS-76 alone
- BLISS-52 plus BLISS-76 populations pooled across trials
- A HDAS (the "Target population") pooled across BLISS-52 and BLISS-76 trials

As results for the Target population in BLISS-52 and BLISS-76 were not supplied separately in the original submission, these results were requested and supplied during the clarification process. Results for Target population patients who received the 1mg/kg regimen are not in the public domain and the manufacturer stated that they were unable to supply these results within the time constraints of the clarification process because of the large amount of other information that was requested.

4.2.2 Description and critique of manufacturers approach to validity assessment

In the main text of the MS, validity assessment of the BLISS trials consisted only of a tabulated quality assessment checklist (MS Table 5.14 page 100); this is reproduced in the two left hand columns in Table 5 below. Further details were provided in MS Appendix 3 (BLISS trials) and in MS Appendix 9 (adverse event studies). The MS was not clear about how their assessment was conducted, or by who, or whether it was based upon the full HGS clinical trial reports or on other information.

A single ERG reviewer undertook an independent quality assessment of the Phase III trials. The MS provided insufficient information for full quality assessment and so additional information in FDA documents[3,5]and in the Navarra publication[14] of the BLISS-52 trial was also utilised. The assessment is summarised in Table 5 below.

Table 5: Quality assessment and ERG critique of BLISS-52 and BLISS-76 trials

Question	MS rating	ERG rating	ERG comment
Was randomisation carried out appropriately?	Yes	Yes	MS states randomisation was stratified and MS and Navarra[14]state that a computer generated randomisation schedule was created.
Was the concealment of treatment allocation adequate?	Yes	Yes	Unmasked pharmacist prepared unmarked treatment infusion bags; but MS not explicit whether the creation and ownership of the randomisation schedule was handled by a separate group who had no direct involvement in the study.
Were the groups similar at the outset of the study in terms of prognostic factors?	Yes	Yes	Agree
Were the care providers, participants and outcome assessors blind to treatment allocation?	Yes	Yes	MS states: *patients, investigators, study coordinators, and sponsors were masked to treatment assignment during intravenous administration of the drug and assessment of the patients every 4 weeks during the 52-week trial*. But methods not described and adherence to blinding not investigated.
Were there any unexpected imbalances in drop-outs between groups?	No	No	Table 9 of the FDA briefing package[3]provides the relevant information
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No	No	The submission did not report outcome results for the 1 mg/kg treatment arms of the trials. For the total population these are available in the public domain.
Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Yes	No	Analysis was done in a modified intention-to-treat (mITT) population of all randomly assigned patients who received a dose of belimumab. The mITT analysis was performed according to the treatment that a subject was randomized to receive, regardless of actual treatment received. This was appropriate.
Adapted from Centre for Reviews and Dissemination (2008) Systematic reviews. CRD's guidance for undertaking reviews in health care. York: Centre for Reviews and Dissemination			

Trial conduct

The BLISS trials were large, international, multi-centre, double-blind, placebo-controlled studies with three parallel groups, that employed a novel primary outcome measure which required proficiency training for assessors. According to the HGS FDA briefing document.[5]

"*they were conducted under Special Protocol Assessment agreement with the FDA with special agreement with respect to selected patient population, primary end point, sample size, stratification factors, statistical methods and concomitant medication controls*".

Randomisation

The MS Appendix 3 provided the following description of randomisation:

"*Patients who underwent all screening procedures and met the entry criteria were enrolled in the study and assigned to treatment by use of a central interactive voice response system. Patients were randomised in a 1:1:1 ratio to placebo, or belimumab 1 mg/kg or 10 mg/kg. Randomisation was stratified according to the SELENA-SLEDAI score (6–9 vs≥10), proteinuria concentration (<2 g/24 h vs≥2 g/24 h) at screening, and ethnic origin (African descent or indigenous American [Alaska Native or American Indian from North, South, or Central America] vs other).*"

ERG note that in the BLISS-52 trial 867 patients were randomised, and that 142 screened-patients who met inclusion criteria were not randomised; the corresponding numbers for BLISS-76 were: 826 randomised and 135 not randomised (data from MS Figures 5.4 and 5.5).

Allocation concealment

MS Appendix 3 states:

"An unmasked pharmacist prepared unmarked infusion bags for administration. Belimumab and placebo were both prepared as sterile and lyophilised vials (5 mL for belimumab 1 mg/kg; 20 mL for belimumab 10 mg/kg and placebo), and contained the same formulations, except without the active drug for placebo."

The ERG considers that the above provides some assurance that allocation concealment was maintained but notes the difficulties of maintaining concealment across large multi-centred studies.

Baseline balance between treatment groups

Data provided in the MS and in FDA documents[3,5] indicates that within each trial there was a reasonable balance between known and putative prognostic factor.

Blinding of treatment allocation

MS Appendix 3 states:

"Patients, investigators, study coordinators, and sponsors were masked to treatment assignment during intravenous administration of the drug and assessment of the patients every 4 weeks during the 52-week trial until the database was locked."

The above gives the manufacturer's description of blinding of care providers, patients and outcome assessors to treatment allocation. There is no mention of methods employed (e.g. all potential flares should be adjudicated by a data monitoring board blinded to treatment). The methods for, extent of, and any problems with, blinding were not described. In the ERG's opinion it is possible that BLISS-52 physician outcomes assessors might have been unblinded, thus explaining a more positive PGA in the intervention group in this study as compared to the PGA in the BLISS-76 study.

Imbalance of drop outs between groups

MS appendices 3 and 7 and state respectively:

"There were no differences among groups in discontinuation rates" "The three groups did not differ in reasons for discontinuation of treatment." and *"Withdrawals and dropouts were adequately reported"*.

Drop outs were reasonably balanced between treatment arms. Infringement of concomitant medication rules was one reason for discontinuation of treatment, and this differed between treatment arms. According the FDA[3] analysis: *" unlike dropouts, 'medication failures' are not balanced across treatment groups (17%, 9%, and 10% for placebo, 1 mg/kg belimumab, and 10 mg/kg belimumab respectively in* BLISS-76 *study, and 11%, 7%, and 6% for* BLISS-52".

The MS did not provide relevant information about adherence of study medication (e.g. missed infusion due to missed clinic visits).

Intention to treat analysis

MS Appendix 3 states:

"Analysis was done in a modified intention-to-treat population, defined as all randomly assigned patients who received a dose of the study drug. This was appropriate and appropriate methods for handling missing data were outlined in the clinical study report."

The trials were analysed according to a modified intention treat (mITT) procedure. In BLISS-52 and in BLISS-76 respectively two of 867 randomised patients and 7 of 826 randomised patients withdrew before receiving medication. Outcome analyses were based on the remaining 865 and 819 patients according to their randomisation group. Thus the results of mITT analyses were unlikely to differ substantially from a full ITT.

Pooling of trials

The pooling of trial data across trials is considered in sections 4.2.4 and 4.2.6 in the current report.

Applicability to the UK and UK clinical practice

It is unclear how many of the 1684 patients recruited to the BLISS-52 and BLISS-76 trials were from UK centres. The ERG notes that patients in the trials were derived from other EU countries. The MS is unclear whether similar care pathways to the UK occur across all centres included in the trials.

4.2.3 Description and critique of manufactures outcome selection

The primary efficacy endpoint in both Phase III studies was the percentage of responders at week 52 estimated using the SRI. The SRI is a novel composite outcome which was developed in consultation with the FDA during protocol planning for the BLISS-52 and BLISS-76 trials. The manufacturer's submission to the FDA states that assessors received proficiency training in SRI outcome assessment at all the centres involved in the trials. SRI and other outcomes selected for reporting in the MS are listed in Table 6. The ERG considers these outcomes to be appropriate for the decision problem.

Table 6: Outcomes reported in MA

Outcome	Measure	Outcome specification
SLE Responder Index (SRI*)	% responders at wk 52	Specified primary outcome
Reduction in SLEDAI score by ≥ 4 points	% responders at wk 52	Specified major secondary outcome
Change in PGA score from baseline	Mean change at wk 24	Specified major secondary outcome
Steroid reduction weeks 40 to 52	% responders	Specified major secondary outcome
SF-36 Physical component summary score	Mean change at wk 24	Specified major secondary outcome
SLE Responder Index	% responders at week 76	Specified major secondary outcome
SLICC/ACR damage index	Mean change at wk 52	Specified secondary outcome
FACIT-fatigue scale mean change from baseline	Mean change at clinic visits	Specified secondary outcome
EQ-5D score	Mean change at clinic visits	Specified secondary outcome
Change in PGA score from baseline	Mean change at wk 52	Specified secondary outcome
SF-36 Physical component summary score	Mean change at wk 52	Specified secondary outcome
SLEDAI SLE flare index over 52 wks	Time to first flare	Specified secondary outcome
SLE Responder Index (SRI)	% responders at timed clinic visits	Other outcome reported
Modified SLE responder index	% responders at wk 52	Other outcome reported
No worsening in PGA score by ≥ 0.3	% responders at wk 52	Other outcome reported
No new BILAG 1A/2B domain scores	% responders at wk 52	Other outcome reported
Change in SLEDAI score from baseline	Mean change at week 52	Other outcome reported

* Composite outcome measure consisting of ≥ 4 points improvement in SLEDAI score, no worsening in PGA by ≥ 0.3 points and no new BILAG 1A or 2B domain scores
FACIT= Functional Assessment of Chronic Illness Therapy
EQ-5D = EuroQoL 5 dimensions
BILAG = British Isles Lupus Assessment Group
SLEDAI = Systemic Lupus Erythematosus Disease Activity Index
SF-36 = Short Form 36-Item Health Survey
SLICC = Systemic Lupus International Collaborating Clinics
ACR = American College of Rheumatology

4.2.4 Description and critique of the statistical approach used

The manufacturer's approach is described in Table 5.13 of the MS (Page 91).

Binary efficacy variables were assessed with a logistic regression model, continuous variables were analysed with an analysis of covariance model, and time-to-flare variables were analysed by use of a Cox proportional hazards model. All analyses were adjusted for baseline randomisation factors. In addition, the JHU observational cohort of patients was used to generate the analysis that was used in an SLE patient simulation.

The ERG reviewed the statistical approach submitted in the main report and notes that in general, the statistical methodologies proposed are suitable to these types of data. However, the ERG identified a number of concerns as shown below:

In order to identify baseline factors that were predictive of response at Week 52 irrespective of treatment received and to evaluate belimumab treatment effect adjusted for the predictive factors, a logistic regression main effects model was developed by the manufacturer based on the pooled data from the Phase III studies (BLISS-52 and BLISS-76).

The ERG notes that while the pooling of the two data sets might be considered appropriate, given that the trials were essentially identical in design and in the analysis of the primary endpoint, the approach used to account for potential between-study variability in the estimate of the baseline response or the uncertainty in the estimate of the population sampling variation was not appropriate (i.e. treatment-by-study interaction). It is not surprising that the P-values for the treatment-by-study interaction were not significant (interaction P-values > 0.5). This insignificant P-value is a reflection of the similarity between the two trials in terms of the primary endpoint and would not capture a real difference that might exist between the two trials.

A mixed model logistic regression would have been appropriate to account for the correlation structure between the two trials and any population sampling variation. Furthermore, a sensitivity analysis of the choice of correlation structure should have been conducted. Without taking into account the unobserved uncertainty or variability between the two trials, the ERG believes that the validity of pooling of data may have been overestimated.

The ERG also note that results of in the manufacturer's submission analysing time-to-flare variables did not take into account the time-varying effects of some of the covariates. A generalized mixed model with time-varying effects could have been considered to deal with the time-varying effect of covariates.

Sub-group analysis

The main submission indicates that a series of pre-specified and post-hoc subgroup analyses for efficacy data were conducted. A comparison between each belimumab treated group and the placebo group was performed by major subgroups which were pre-specified in each Phase III analytical plan.

With reference to the decision problem and the manufacturer's intention to explicitly identify patients who benefit the most, the ERG notes that some additional exploratory subgroup analyses which were not pre-specified in the individual analytical plans were evaluated using the pooled Phase III population Target or high disease activity sub-group. The subgroup of patients with evidence for serological disease activity (low complement and positive anti-dsDNA) and who additionally have a SELENA-SLEDAI disease activity score ≥ 10 at baseline. However, even though patients in this subgroup experienced the greatest treatment effect over and above the total pooled population, the ERG notes that this sub-group analysis was not pre-specified in the analytical plan. Therefore, the results of this sub-group analysis should not be regarded as definitive since the two trials were not powered to conduct this sub-group analysis.

The ERG notes that there was no attempt to summarise the studies by performing a meta-analysis or by conducting an incremental analysis.

4.2.5 Results from pivotal trials

The clinical effectiveness results in the MS are derived from the two BLISS trials.

4.2.5.1 BLISS trial study design and patient eligibility

Methodological details of the BLISS-52 and BLISS-76 trials were presented in Table 5.6 of the MS which is reproduced in Table 7.

Table 7: MS summary of BLISS trial methodology (from MS Table 5.6)

Trial no. (acronym)	**C1057 (BLISS-52)**	**C1056 (BLISS-76)**
Location	90 centres in 13 countries in Latin America (Argentina, Brazil, Chile, Colombia and Peru), Asia-Pacific (Australia, Hong Kong, India, Korea, Philippines and Taiwan) and eastern Europe (Romania and Russia).	136 centres in 19 countries in North America (Canada, Costa Rica, Mexico, Puerto Rico and US) and Europe (Austria, Belgium, Czech Republic, France, Germany, Israel, Italy, The Netherlands, Poland, Romania, Slovakia, Spain, Sweden and UK).
Design	Randomised, double-blind, placebo-controlled, parallel-group study.	As per BLISS-52.
Duration of study	52 weeks	76 weeks (primary end point at 52 weeks)
Method of randomisation	Patients who underwent all screening procedures and met the entry criteria were enrolled in the study and assigned to treatment by use of a central interactive voice response system. Patients were randomised in a 1:1:1 ratio to placebo, or belimumab 1 mg/kg or 10 mg/kg. Randomisation was stratified according to the SELENA-SLEDAI score (6–9 vs ≥10), proteinuria concentration (<2 g/24 h vs ≥2 g/24 h) at screening, and ethnic origin (African descent or indigenous American [Alaska Native or American Indian from North, South, or Central America] vs other).	As per BLISS-52.
Method of blinding (care provider, patient and outcome assessor)	Patients, investigators, study coordinators, and sponsors were masked to treatment assignment during intravenous administration of the drug and assessment of the patients every 4 weeks during the trial until the database was locked. An unmasked pharmacist prepared unmarked infusion bags for administration. Belimumab and placebo were both prepared as sterile and lyophilised vials (5 mL for belimumab 1 mg/kg; 20 mL for belimumab 10 mg/kg and placebo), and contained the same formulations, except without the active drug for placebo.	As per BLISS-52.
Intervention(s) (n=) and comparator(s) (n=)	Standard of care plus belimumab 1mg/kg (n=288) or belimumab 10mg/kg (n=290) or placebo (n=287) administered by IV infusion on Days 0, 14 and 28 and every 28 days thereafter for 48 weeks. Standard of care consisted of the following (alone or in combination): antimalarials, NSAIDs, corticosteroids or other immunosuppressants (azathioprine, methotrexate, and mycophenolate mofetil).	Standard of care plus belimumab 1mg/kg (n=271) or belimumab 10mg/kg (n=273) or placebo (n=275) administered by IV infusion on Days 0, 14 and 28 and every 28 days thereafter for 72 weeks. Standard of care consisted of the following (alone or in combination): antimalarials, NSAIDs, corticosteroids or other immunosuppressants (azathioprine, methotrexate, and mycophenolate mofetil).
Progressive restrictions placed on standard of care	In both BLISS-52 and BLISS-76, progressive restrictions were placed on standard of care as the study progressed. These are outlined in the Figure 5.2 below. (see following box)	
Primary outcomes (including scoring methods and timings of assessments)	The primary efficacy endpoint was the response rate at week 52, assessed with SLE Responder Index (SRI). With the SRI criteria, a responder was defined as having a reduction of at least 4 points in the SELENA-SLEDAI score (defined as clinically meaningful) [15], no new BILAG A organ domain score, no more than 1 new BILAG B organ domain score, and no worsening in PGA score (increase <0.3) at week 52 compared with baseline.	As per BLISS-52.
Secondary outcomes (including scoring methods and timings of assessments)	Major secondary endpoints: • Percent of subjects with ≥ 4-point reduction in SELENA-SLEDAI at Week 52. • Mean change in PGA at Week 24. • Percent of subjects with prednisone (equivalent) reduction ≥ 25% from baseline to ≤ 7.5 mg/day during Weeks 40 – 52 (in subjects whose prednisone equivalent dose was > 7.5 mg/day at baseline). • Mean change in SF-36 PCS at Week 24.	Major secondary endpoints: • As per BLISS-52. • Additionally, response rate (SRI) at Week 76.
Duration of follow-up	52 or 56 weeks dependent on participation in the continuation protocol.	76 or 80 weeks dependent on participation in the continuation protocol.

COMMENT

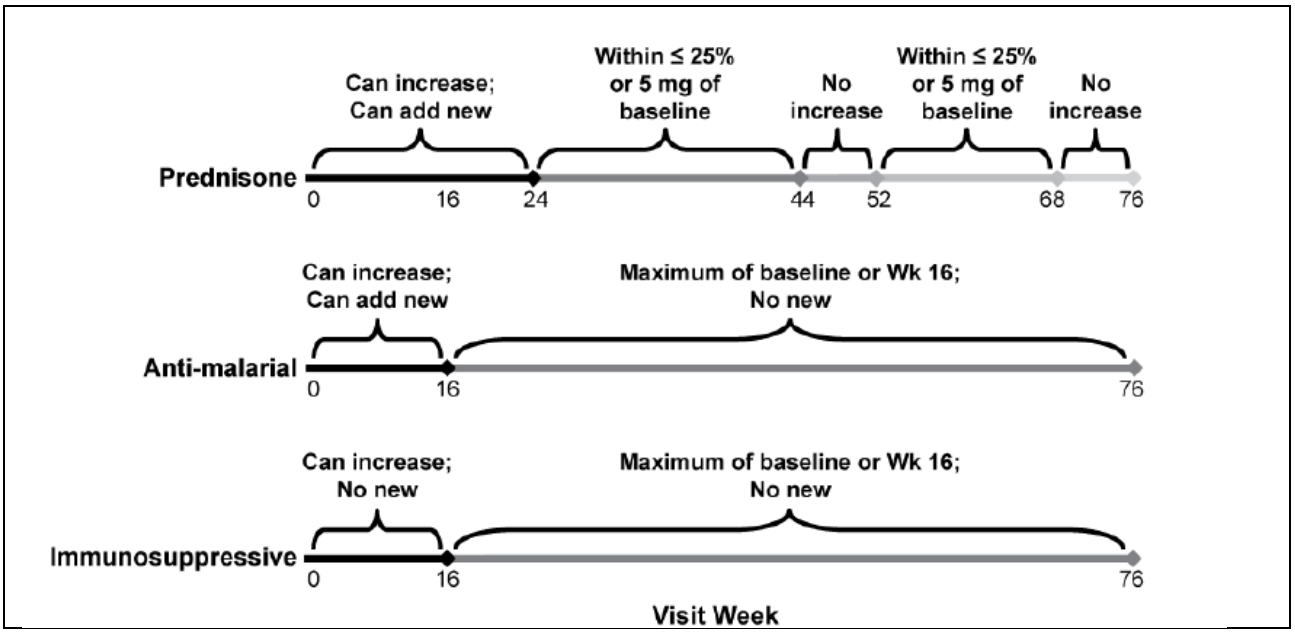
The two pivotal trials, BLISS-52 and BLISS-76, were international multicentre RCTs undertaken in different largely non-overlapping geographical regions. The geographical location of study centres differed considerably between trials. In BLISS-52 there were 90

centres: in 13 countries in Latin America there were 38 centres (Argentina, Brazil, Chile, Colombia and Peru), in Asia-Pacific there were 41 centres (Australia, Hong Kong, India, Korea, Philippines and Taiwan) and in Eastern Europe there were eleven centres (Romania and Russia). In BLISS-76 there were 136 centres in 19 countries in North America (Canada, Costa Rica, Mexico, Puerto Rico and US) and Europe (Austria, Belgium, Czech Republic, France, Germany, Israel, Italy, The Netherlands, Poland, Romania, Slovakia, Spain, Sweden and UK). North America (65 centres) and Europe (62 centres) contributed 93% of the centres in BLISS-76. These geographical differences were reflected in racial differences between the populations in the two trials. Although both trials included adults with auto-antibody positive active SLE it is arguable that the population in BLISS-76 is more likely to be similar to that in England and Wales than that from BLISS 52. It is reasonable to assume that the results from BLISS-76 are more generalisable to the UK.

Randomisation was stratified according to SELENA-SLEDAI score (6–9 vs ≥10), proteinuria concentration (<2 g/24 h vs ≥2 g/24 h) at screening, and ethnic origin.

Progressive constraints on standard care medications (immunosuppressives, anti-malarials and steroids) (see Figure 1) were imposed during the trials; these were implemented so as to increase the possibility of detecting improvement due to belimumab without interference from the effects of changing background standard care treatments.

Figure 1: Constraints on standard of care medications (MS Figure 5.2)



Patient eligibility for BLISS-52 and BLISS-76 was the same and summarised below in Table 8.

Table 8: Patient eligibility for BLISS-52 and BLISS-76 (From MS Table 5.7)

Trial no. (acronym)	Inclusion criteria	Exclusion criteria
C1057 **(BLISS-52)**	Adult patients (aged ≥ 18 years) who met the American College of Rheumatology criteria for systemic lupus erythematosus and had active disease (score ≥ 6 at screening on SELENA-SLEDAI) were eligible for enrolment. Other inclusion criteria were unequivocally positive ANA (titre ≥ 1:80) or anti-dsDNA antibody (≥ 30IU/mL), and a stable treatment regimen with fixed doses of prednisone (0–40mg/day), or non-steroidal anti-inflammatory, antimalarial, or immunosuppressant drugs for at least 30 days before the first study dose	The main exclusion criteria were severe active lupus nephritis or CNS lupus; pregnancy; and previous treatment with any B-lymphocyte-targeted drug (including rituximab), intravenous cyclophosphamide within 6 months of enrolment, and intravenous Ig or prednisone (>100 mg/day) within 3 months
C1056 **(BLISS-76)**	As per BLISS-52	As per BLISS-52
Adapted from Pharmaceutical Benefits Advisory Committee (2008) Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee (Version 4.3). Canberra: Pharmaceutical Benefits Advisory Committee		

Relative to the whole trial population imbalance between treatment arms was more pronounced for the Target population in both trials, especially in BLISS-76 (see Appendix 4 of this report).

4.2.5.2 BLISS trials: demography of patients

Demographic characteristics of patients in BLISS-52 and BLISS-76 were presented in MS Tables 5.8 to 5.11 (see Appendix 4 of this report). Patients were mostly young females (74% ≤ 45 years of age; 94% female), a population which is representative of patients with SLE.

Selected characteristics for placebo and 10mg/kg groups taken from MS Table 5.8 are shown below in Table 9. Amongst all treatment arms pooled across the two studies 47% of patients were white, 23% American Indian, 21% Asian, and 8.8% black, however there were large differences in the racial makeup between the two studies reflecting the racial distributions in the geographic regions in which the trial centres were located. The substantial differences between trials in geographical and in racial distributions seen for the whole population were mirrored in the Target population Table 9 (Appendix 4).

Table 9: Demographic characteristics in the BLISS trials (adapted from MS Table 5.8)

Table 5.8. Selected demographic characteristics in Phase 3 trials

	BLISS-52			BLISS-76			Pooled Total Population		
Race[1]	**Placebo N = 287**	**10mg/kg N = 290**	**All N = 865**	**Placebo N = 275**	**10mg/kg N = 273**	**All N = 819**	**Placebo N = 562**	**10mg/kg N = 563**	**All N = 1684**
White	82 (28.6%)	71 (24.5%)	229 (26.5%)	188 (68.4%)	189 (69.2%)	569 (69.5)	270 (48.0%)	260 (46.2%)	798 (47.4%)
Asian	105 (36.6%)	116 (40.0%)	327 (37.8%)	11 (4.0%)	11 (4.0%)	28 (3.4%)	116 (20.6%)	127 (22.6%)	355 (21.1%)
Black	11 (3.8%)	11 (3.8%)	30 (3.5%)	39 (14.2%)	39 (14.3%)	118 (14.4%)	50 (8.9%)	50 (8.9%)	148 (8.8%)
Alaska Native or American Indian from North/Central/ South America	89 (31.0%)	92 (31.7%)	279 (32.3%)	36 (13.1%)	34 (12.5%)	103 (12.6%)	125 (22.2%)	126 (22.4%)	382 (22.7%)
Hispanic or Latino origin	143 (49.8%)	136 (46.9%)	420 (48.6%)	55 (20.0%)	56 (20.5%)	173 (21.1%)	198 (35.2%)	192 (34.1%)	593 (35.2%)

[1] Patients who checked more than 1 race category are counted under individual race category according to the minority rule as well as the multiracial category.

Differences also existed between studies in that BLISS-76 patients had longer disease duration and more organ damage (higher SLICC damage scores), and were using lower steroid dosages than BLISS-52 patients.

Both BLISS-52 and -76 populations presented a restricted range of SLE manifestations. The MS did not provide tabulated information for the frequency of SELENA SLEDAI manifestations at baseline, these are shown in Table 10 below based on the FDA discussion document3 for the whole BLISS populations, and in Table 11 for the target population. The majority of BLISS-76 participants had musculoskeletal and/or mucocutaneous manifestations of SLE as assessed by the SELENA SLEDAI disease activity index. Baseline disease involvement was generally well balanced within trial between the three treatment groups with the exception of rash. Higher proportions of placebo patients (68%) and patients in the 1mg/kg belimumab group (66%) had a rash at study entry as compare to patients in the 10mg/kg (56%). A similar pattern of SLE disease involvement at baseline was observed for subjects in BLISS-52, however, a lower rate of arthritis (59%) was reported compared to BLISS-76 (72%).

Table 10: Baseline SELENA SLEDAI involvement: whole population in BLISS trials

	BLISS-52				BLISS-76			
Condition (weight)	**Placebo (N=287)**	**Belimumab 1mg/kg (N=288)**	**Belimumab 10mg/kg (N=290)**	**Total (N=865)**	**Placebo (N=275)**	**Belimumab 1mg/kg (N=271)**	**Belimumab 10mg/kg (N=273)**	**Total (N=819)**
Organic Brain Syndrome (8)	0	2 (1%)	0	2 (1%)	1 (0%)	2 (1%)	3 (1%)	6 (1%)
Lupus HA (8)	4 (1%)	2 (1%)	4 (1%)	10(1%)	1 (0%)	4 (2%)	9 (3%)	14 (2%)
Vasculitis (8)	20 (7%)	16 (6%)	28 (10%)	64 (7%)	17 (6%)	20 (7%)	10 (4%)	47 (6%)
Arthritis (4)	165 (58%)	169 (59%)	173 (60%)	507 (59%)	206 (75%	193 (71%)	191 (70%	590 (72%)
Hematuria (4)	15 (5%)	16 (6%)	16 (6%)	47(5%)	5 (2%)	7 (3%)	8 (3%)	20 (2%)
Proteinuria (4)	50 (19%)	54(19%)	41 (14%)	145 (17%)	29 (11%)	23 (9%)	26 (10%)	78 (10%)
Rash (2)	176 (61%)	176 (61%)	182 (63%)	534 (62%)	87 (68%)	180 (66%)	154 (56%)	521 (64%)
Alopecia (2)	150 (52%)	138 (48%)	158 (55%)	446 (52%)	30 (47%	137 (51%)	116 (43%	383 (47%)
Mucosal Ulcers (2)	71 (25%)	52 (18%)	58 (20%)	181 (21%)	74 (27%)	57 (21%)	78 (29%)	209 (26%)
Low Complement (2)	183 (64%)	186 (65%)	198 (68%)	567 (66%)	160 (58%	149 (55%)	159 (58%	468 (57%)
Inc. DNA Binding (2)	205 (71%)	220 (76%)	218 (75%)	643 (74%)	175 (64%	168 (62%)	176 (65%	519 (63%)
Leukopenia (1)	18 (6%)	12 (4%)	9 (3%)	39 (5%)	16 (6%)	22 (8%)	23 (8%)	61 (7%)

Table 11: Baseline SELENA SLEDAI involvement: in the Target population in BLISS Trials

	BLISS-52		BLISS-76		Combined BLISS	
Condition (weight)	**Placebo (N=107)**	**Belimumab 10mg/kg (N=112)**	**Placebo (N=96)**	**Belimumab 10mg/kg (N=96)**	**Placebo (N=203)**	**Belimumab 10mg/kg (N=193)**
Organic Brain Syndrome (8)	0	0	1 (1.0%)	0	1 (0.5%)	0
Lupus HA (8)	1 (0.9%)	3 (2.7%)	0	2 (2.5%)	1 (0.5%)	5 (2.6%)
Vasculitis (8)	15 (14.0%)	19 (17.0%)	10 (10.4%)	5 (6.2%)	25 (12.3%)	24 (12.4%)
Arthritis (4)	65 (60.7%)	76 (67.9%)	83 (86.5%)	63 (77.8%)	148 (72.9%)	139 (72.0%)
Hematuria (4)	9 (8.4%)	7 (6.3%)	3 (3.1%)	6 (7.4%)	12 (5.9%)	13 (6.7%)
Proteinuria (4)	31 (29.0%)	28 (25.0%)	17 (17.7%)	21 (25.9%)	48 (23.6%)	49 (25.4%)
Rash (2)	74 (69.2%)	75 (67.0%)	72 (75.0%)	52 (64.2%)	146 (71.9%)	127 (65.8%)
Alopecia (2)	66 (61.7%)	69 (61.6%)	50 (52.1%)	38 (46.9%)	116 (57.1%)	107 (55.4%)
Mucosal Ulcers (2)	28 (26.2%)	20 (17.9%)	30 (31.3%)	22 (27.2%)	58 (28.6%)	42 (21.8%)
Low Complement (2)	107 (100.0%)	112 (100.0%)	96 (100.0%)	80 (98.8%)*	203 (100.0%)	192 (99.5%)
Inc. DNA Binding (2)	107 (100.0%)	112 (100.0%)	96 (100.0%)	81 (100.0%)	203 (100.0%)	193 (100.0%)
Leukopenia (1)	6 (5.6%)	4 (3.6%)	7 (7.3%)	10 (12.3%)	13 (6.4%)	14 (7.3%)

A specified major secondary outcome was the percentage of SRI responders at week 76. There was only a small difference between placebo and 10mg/kg belimumab (odds ratio and P value not submitted; odds ratio 1.31, 95% CI: 0.92 – 1.87, P = 0.1323 by logistic regression, taken from the FDA HGS briefing document.[5]

Relative to the whole trial population imbalance between treatment arms was more pronounced for the Target population in both trials, especially in BLISS-76 (Appendix 4).

Patients from BLISS-52 contributed more patients to the pooled Target population than did patients from BLISS-76 (55% and 45% respectively, and contributed a greater proportion of the patients receiving 10mg/kg belimumab (58% and 42% from each trial respectively); therefore effectiveness results pooled across trials will tend to reflect BLISS-52 outcomes more than BLISS-76.

4.2.5.3 BLISS trial results by outcome

Primary outcome: SRI at week 52

The pre-specified primary outcome in the BLISS trials was the proportion of responders at week 52 defined according to the composite SRI outcome measure. The results were provided in MS Table 5.15 and clarification Table A6.1 and summarised below in Table: 12.

Table: 12 Primary efficacy endpoint (SRI) at Week 52 (dropout-failure)

	BLISS-52		**BLISS-76**		**Pooled Total Population[4]**		**High Disease Activity Subgroup Pooled Total**		**High Disease Activity Subgroup BLISS-52**		**High Disease Activity Subgroup BLISS-76**	
SRI at Week 52	**Placebo N = 287**	**10mg/kg N = 290**	**Placebo N = 275**	**10mg/kg N = 273**	**Placebo N = 562**	**10mg/kg N = 563**	**Placebo N = 203**	**10mg/kg N = 193**	**Placebo N = 107**	**10mg/kg N = 112**	**Placebo N = 96**	**10 mg/kg N = 81**
No. (%) Response	125 (43.6%)	167 (57.6%)	93 (33.8%)	118 (43.2%)	218 (38.8%)	285 (50.6%)	77 (37.9%)	121 (62.7%)	44 (41.1%)	75 (67.0%)	33 (34.4%)	46 (56.8%)
Observed difference vs placebo (%)	-	14.03	-	9.41	-	11.8	-	24.8	-	25.9	-	22.4
OR (95% CI)[1] vs placebo	-	1.83 (1.30, 2.59)	-	1.52 (1.07, 2.15)	-	1.68 (1.3, 2.2)	-	2.7 (1.8, 4.1)	-	3.0 (1.7, 5.2)	-	2.5 (1.3, 4.6)
P-value[1]	-	0.0006	-	0.0207	-	< 0.0001	-	< 0.0001		0.0001	-	0.0045

[1] Odds Ratio (95% confidence interval) and p-values were from logistic regression for the comparison between each belimumab dose and placebo with covariates. For individual studies, covariates include baseline SELENA-SLEDAI score (≤ 9 vs ≥ 10), baseline proteinuria level (< 2 g/24 hour vs ≥ 2 g/24 hour equivalent) and race (African descent or indigenous-American descent vs other). For pooled data analysis, study was also included as an additional covariate.

In both trials SoC + 10mg/kg belimumab delivered a greater percentage of responders than SoC + placebo. The difference in percentage of responders in the belimumab group relative to placebo group for the whole population was 14% in BLISS-52 and 9.4% in BLISS-76. The corresponding adjusted odds ratios for a response in BLISS-52 and in BLISS-76 were respectively 1.83 (95% CI: 1.30, 2.59; P = 0.0006) and 1.52 (95% CI: 1.07, 2.15; P = 0.027). For the Target population pooled across trials the difference in percentage of responders in the belimumab group relative to placebo group was 24.8% and the adjusted odds ratio was 2.7 (95% CI: 1.8, 4.1; P < 0.0001). In BLISS-52 and BLISS-76 Target populations the difference

between groups was 25.9% and 22.4% respectively (odds ratio 3.0, 95% CI: 1.7, 5.2; P = 0.0001 for BLISS-52 and odds ratio 2.5, 95% CI: 1.3, 4.6; P = 0.0045 for BLISS-76).

Relative to the whole population the Target population generated results that were more supportive of belimumab. For the whole population and for the Target population BLISS-52 produced results more supportive of belimumab than did BLISS-76, however for the Target population the difference between trials was less than for the total population.

SRI at successive clinic visits and at week 76

The percentage of SRI responders was also reported at multiple follow up times (MS Figures 5.6, to 5.9 shown in Figure 2.

Figure 2: Percentage of SRI responders during follow up (from MS Figures 5.6 to 5.9)

For the Target population pooled across trials and in BLISS-52, at many times, a significantly greater response was observed for the belimumab group relative to placebo group

(significance tests uncorrected for multiple testing), however, for BLISS-76 the only time a significantly (P < 0.05) greater response was observed for the belimumab group was at week 52.

In the HGS/FDA[5]analysis there is little difference in response between 1mg/kg and 10mg/kg groups for BLISS-52.

A specified major secondary outcome was the percentage of SRI responders at week 76. There was only a small difference between placebo and 10mg/kg belimumab (odds ratio and P value not submitted; odds ratio 1.31, 95% CI: 0.92 – 1.87, P = 0.1323 by logistic regression, taken from the FDA discussion document[5]).

The HGS Briefing Document to the FDA[5] provided graphs for all three randomised groups (placebo, 1mg/kg belimumab and 10mg/kg belimumab) for the percentage of SRI responders observed at successive clinic visits up to 52 weeks for BLISS-52 and week 76 for BLISS-76. These graphs are in Figure 3. They indicate that in BLISS-76 there was a minimal difference in response between 1mg/kg and 10mg/kg groups. Baseline characteristics for the three groups (HGS Briefing Document Pages 87 to 100[5]) do not provide an obvious explanation for this result.

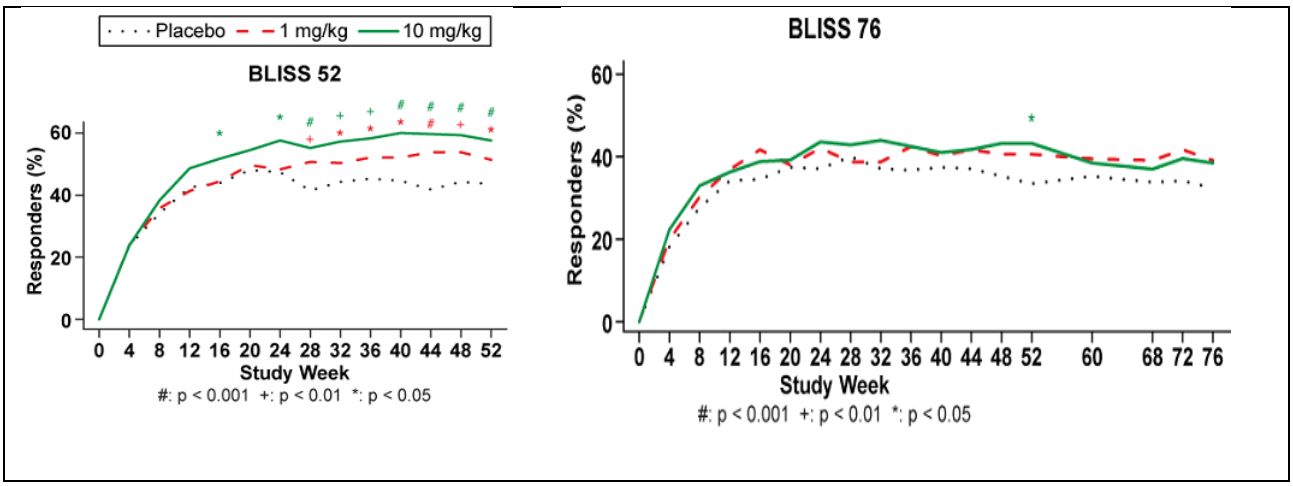


Figure 3: SRI percent responders over follow up (from HGS Briefing Document to FDA)

The ERG note that the percentage SRI responders observed at various follow up times is a group response and does not reflect sustained SRI response at the individual level. The graph line showing percentage of responders across the duration of the trials rose and fell at various follow up times, thus an individual non-responder could later improve sufficiently to be classified as a responder.

Modified SRI

To be classified as an SRI responder a patient is required to have a SELENA-SLEDAI score that is reduced by≥ 4 points relative to baseline. A 4 point reduction in SELENA -SLEDAI can be achieved by normalisation of serological manifestations only (e.g. anti-dsDNA antibodies and complement). The MS presented an analysis of a modified SRI response in which the increased DNA binding and low complement items were removed from the SELENA-SLEDAI component of the SRI; the analysis was performed in patients who still had a SELENA SLEDAI score≥ 4 at baseline after points for low complement and increased DNA binding were removed from the scale. During the clarification process the manufacturer provided modified SRI results for the Target or high disease activity population; these plus the information from the MS Page 111 are summarised in Table 13.

Table 13: Modified SRI response at week 52

	BLISS-523		**BLISS-763**		**Pooled Total Population2**		**High Disease Activity Subgroup Pooled Total**		**High Disease Activity Subgroup BLISS-52**		**High Disease Activity Subgroup BLISS-76**	
Change from Baseline at Week 522	**Placebo N = 264**	**10 mg/kg N = 259**	**Placebo N = 255**	**10 mg/kg N = 245**	**Placebo N = 519**	**10 mg/kg N = 504**	**Placebo N = 203**	**10 mg/kg N = 193**	**Placebo N = 107**	**10 mg/kg N = 112**	**Placebo N = 96**	**10 mg/kg N = 81**
n(%) responders	127 (48.1%)	158 (61.0%)	92 (36.1%)	109 (44.5%)	219 (42.2%)	267 (53.0%)	42 (39.3%)	73 (65.2%)	29 (30.2%)	43 (53.1%)	71 (35.0%)	116 (60.1%)
OR (95% CI)[1]	-	-	-	-	-	-	-	3.0 (1.7, 5.2)	-	2.5 (1.4, 4.8)	-	2.8 (1.8, 4.2)
P-value1	-	-	-	-	-	-	-	0.0001	-	0.0036	-	<0.0001
10 mg/kg vs placebo difference	-	(11.9%)	-	(8.4%)	-	(10.8%)	-	(25.9%)	-	(25.9%)	-	(25.1%)
P-value	-	0.0038	-	0.0604	-	0.0006	-	-	-	-	-	-

[1] ANCOVA model for the comparison between each belimumab dose and placebo, adjusted for baseline SELENA SLEDAI score (≤ 9 vs. ≥ 10), baseline proteinuria level (< 2 g/24 hour vs. ≥ 2 g/24 hour equivalent) and race (African descent or indigenous-American descent vs. other). For pooled data analysis, study was also included as an additional covariate

[2] Defined as SRI response with serology components (increased DNA binding and low complement items) removed

[3] Information extracted from HGS Briefing Document to FDA Figure 9.51

Table 14: Modified SRI response at week 52

	BLISS-52[3]		BLISS-76[3]		Pooled Total Population[2]		High Disease Activity Subgroup Pooled Total		High Disease Activity Subgroup BLISS-52		High Disease Activity Subgroup BLISS-76	
Change from Baseline at Week 52[2]	**Placebo N = 264**	**10 mg/kg N = 259**	**Placebo N = 255**	**10 mg/kg N = 245**	**Placebo N = 519**	**10 mg/kg N = 504**	**Placebo N = 203**	**10 mg/kg N = 193**	**Placebo N = 107**	**10 mg/kg N = 112**	**Placebo N = 96**	**10 mg/kg N = 81**
n(%) responders	127 (48.1%)	158 (61.0%)	92 (36.1%)	109 (44.5%)	219 (42.2%)	267 (53.0%)	42 (39.3%)	73 (65.2%)	29 (30.2%)	43 (53.1%)	71 (35.0%)	116 (60.1%)
OR (95% CI)[1]	-	-	-	-	-	-	-	3.0 (1.7, 5.2)	-	2.5 (1.4, 4.8)	-	2.8 (1.8, 4.2)
P-value[1]	-	-	-	-	-	-	-	0.0001	-	0.0036	-	<0.0001
10 mg/kg vs placebo difference	-	(11.9%)	-	(8.4%)	-	(10.8%)	-	(25.9%)	-	(25.9%)	-	(25.1%)
P-value	-	0.0038	-	0.0604	-	0.0006	-	-	-	-	-	-

[1] ANCOVA model for the comparison between each belimumab dose and placebo, adjusted for baseline SELENA SLEDAI score (≤ 9 vs. ≥ 10), baseline proteinuria level (< 2 g/24 hour vs. ≥ 2 g/24 hour equivalent) and race (African descent or indigenous-American descent vs. other). For pooled data analysis, study was also included as an additional covariate
[2] Defined as SRI response with serology components (increased DNA binding and low complement items) removed
[3] Information extracted from HGS Briefing Document to FDA Figure 9.51

The MS did not specify patient numbers for this analysis and so data from the HGS Briefing Document to the FDA.[5] Figure 4 shows the percentage of modified SR responders (from HGS Briefing Document to FDA).

In the HGS/FDA[5] analysis there is little difference in response between 1mg/kg and 10mg/kg groups for BLISS-52.

The number of patients at risk was not specified. A stronger response was observed for the Target populations than for the total populations and statistical significance was reached in both trials.

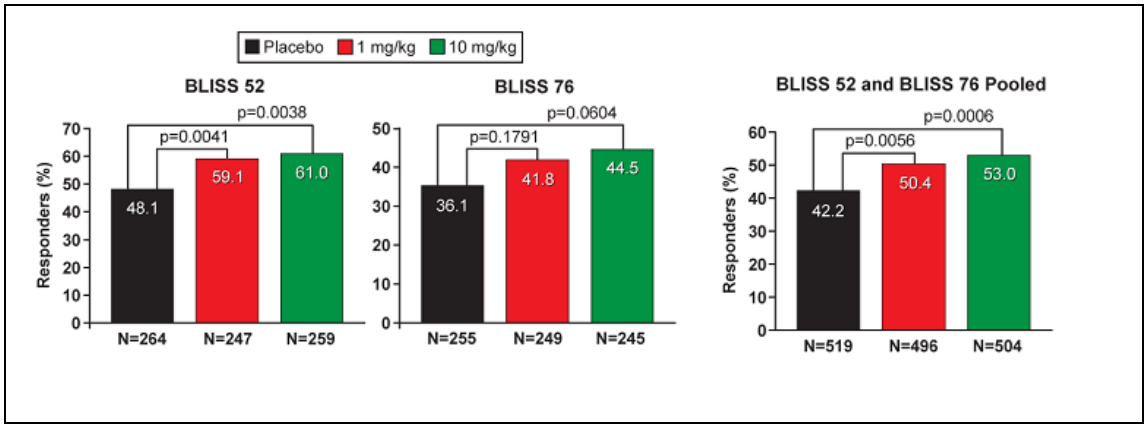


Figure 4: Modified SR percentage of responders (from HGS Briefing Document to FDA)

Subcomponents of the SRI response

Table: 15 summarises the week 52 results for the three subcomponents of the composite SRI response (based on MS Table 5.16 and clarification Table A6.1).

Table 15: Results for subcomponents of SRI at week 52 (adjusted)

	BLISS-52		BLISS-76		Pooled Total Population[4]		High Disease Activity Subgroup Pooled Total		High Disease Activity Subgroup BLISS-52		High Disease Activity Subgroup BLISS-76	
	Placebo N = 287	10mg/kg N = 290	Placebo N = 275	10mg/kg N = 273	Placebo N = 562	10mg/kg N = 563	Placebo N = 203	10mg/kg N = 193	Placebo N = 107	10mg/kg N = 112	Placebo N = 96	10mg/kg N = 81
4-point reduction in SELENA-SLEDAI	132 (46.0%)	169 (58.3%)	98 (35.6%)	128 (46.9%)	230 (40.9%)	297 (52.8%)	84 (41.4%)	125 (64.8%)	47 (43.9%)	76 (67.9%)	37 (38.5%)	49 (60.5%)
Observed difference vs placebo (%)	-	12.3	-	11.3	-	11.9	-	23.4	-	24.0		22.0
OR (95% CI)[1] vs placebo	-	1.71 (1.21,2.41)	-	1.63 (1.15,2.32)	-	1.68 (1.3,2.2)	-	2.6 (1.7,3.9)	-	2.8 (1.6,4.8)	-	2.4 (1.3,4.4)
P-value[1]		0.0024		0.0062	-	< 0.0001	-	< 0.0001		0.0004	-	0.0063
No New 1A/2B BILAG domain scores	210 (73.2%)	236 (81.4%)	179 (65.1%)	189 (69.2%)	389 (69.2%)	425 (75.5%)	125 (61.6%)	145 (75.1%)	68 (63.6%)	88 (78.6%)	57 (59.4%)	57 (70.4%)
Observed difference vs placebo (%)	-	8.2	-	4.1	-	6.3	-	13.6	-	15.0	-	11.0
OR (95% CI)[1,2] vs placebo	-	1.62 (1.09,2.42)	-	1.20 (0.84,1.73)	-	1.4 (1.1,1.8)	-	1.9 (1.2,3.0)	-	2.3 (1.2,4.2)	-	1.6 (0.9, 3.1)
P-value[1,2]		0.0181		0.3193	-	0.0190	-	0.0034	-	0.0099	-	0.1297
No worsening in PGA	199 (69.3%)	231 (79.7%)	173 (62.9%)	189 (69.2%)	372 (66.2%)	420 (74.6%)	119 (58.6%)	142 (73.6%)	64 (59.8%)	86 (76.8%)	55 (57.3%)	56 (69.1%)
Observed difference vs placebo (%)	-	10.4	-	6.3	-	8.4	-	15.0	-	17.0	-	11.8
OR (95% CI)[1,3] vs placebo	-	1.74 (1.18,2.55)	-	1.32 (0.92,1.90)	-	1.5 (1.2,2.0)	-	2.0 (1.3,3.1)	-	2.3 (1.3,4.2)	-	1.6 (0.9,3.0)
P-value[1,3]	-	0.0048	-	0.1258	-	0.0017	-	0.0015	-	0.0063	-	0.1312

[1] Odds Ratio (95% confidence interval) and p-value were from logistic regression for the comparison between each belimumab dose and placebo with covariates. For individual studies, covariates include baseline SELENA-SLEDAI score (≤ 9 vs ≥ 10), baseline proteinuria level (< 2 g/24 hour vs≥ 2 g/24 hour equivalent) and race (African descent or indigenous-American descent vs other). For pooled data analysis, study was also included as an additional covariate

[2] Additional covariate: baseline BILAG domain involvement (at least 1A/2B)

[3] Additional covariate: baseline PGA score

[4] No significant treatment-by-study interactions were observed (all p > 0.287)

The three subcomponents of the composite SRI outcome were: [i] an improved SELENA SLE DAI score by≥ 4 points; [ii] a BILAG index showing no new grade A organ involvement or no two grade B organ involvements (i.e. no worsening by one new A or two new B BILAG indices); [iii] a PGA score that has not increased by more than 0.3 points (i.e. no worsening in PGA by ≥ 0.3).

The percentage of patients at week 52 that achieved a SLEDAI score reduction of ≥ 4 points was defined as a major secondary outcome. For the whole population, both trials delivered more responders in the belimumab group than the placebo group (P = 0.0024 and P = 0.0062 for BLISS-52 and BLISS-76, respectively).

Results at week 52 for the other two SRI subcomponents (i.e. no worsening in BILAG index and no worsening in PGA score) were defined as non-major secondary outcomes. The percentage of patients in the whole population that satisfied the BILAG and PGA criteria in BLISS-52 was greater for belimumab relative than placebo (significant at P = 0.0181 and P = 0.0048 for BILAG and PGA, respectively); however, for BLISS-76 the differences between belimumab and placebo were considerably smaller and neither component reached statistical significance in favour of belimumab (P = 0.319 and P = 0.1258 for BILAG and PGA, respectively). According to results reported in the HGS Briefing Document to the FDA (Table 9.20, Page 102) the 1mg/kg belimumab dose regimen in BLISS-76 performed slightly better than 10mg/kg for both the PGA and BILAG subcomponents at week 52.

The corresponding results for the target population supplied during the clarification process are also summarised in Table 15 Pooled across trials, all three SRI components at week 52 were supportive of belimumab relative to placebo and delivered significant effects. However, for BLISS-76 the PGA and BILAG results at week 52 for the target population were considerably weaker (P = 0.1312 and P = 0.1297, respectively) than for BLISS-52 or the pooled target population.

Major secondary outcomes

The MS identified five pre-specified major secondary outcomes. These included the SRI response at week 76 and the percentage of patients with a≥ 4 point SLEDAI improvement at week 52, each of which have been discussed in the preceding sections. The other three major secondary outcomes were: mean change in PGA score at week 24, percentage of patients with prednisone reductions ≥ 25% from baseline to ≤ 7.5 mg/day during weeks 40 to 52 (in subjects whose baseline dose was > 7.5 mg/day); mean change in SF36 PCS at week 24. These are discussed in this section.

Change in PGA score at week 24 was presented in MS Table 5.18 and the relevant results from this are shown in Table: 16 below. For the whole population in BLISS-52 the change in PGA score (week 24 relative to baseline) for both groups indicated disease improvement and was greater in the belimumab group (-0.54) than placebo group (-0.39; P = 0.0003 in support of belimumab). For BLISS-76 the difference between groups was very small and in favour of placebo (-0.49 placebo and -0.48 belimumab) and did not reach statistical significance (P = 0.7987). For the Target HDAP pooled across trials belimumab delivered a greater reduction in PGA score than placebo (P = 0.028 with mean changes of -0.42 and -0.52 for placebo and belimumab, respectively). Target population results by trial are not available.

Table 16: Mean change in PGA score at week 24 (taken from MS Table 5.18)

Major secondary endpoint at Week 24	BLISS-52		BLISS-76		Pooled Total Population		High Disease Activity Subgroup	
	Placebo N = 287	10mg/kg N = 290	Placebo N = 275	10mg/kg N = 273	Placebo N = 562	10mg/kg N = 563	Placebo N = 203	10mg/kg N = 193
Mean ± SE	-0.39 ± 0.03	-0.54 ± 0.03	-0.49 ± 0.04	-0.44 ± 0.03	-0.44 ± 0.02	-0.49 ± 0.02	-0.42 ± 0.04	-0.52 ± 0.04
LS Mean ± SE[1]	-0.35 ± 0.04	-0.50 ± 0.04	-0.49 ± 0.05	-0.48 ± 0.05	-0.40 ± 0.03	-0.48 ± 0.03	-0.41 ± 0.05	-0.53 ± 0.05
P-value[1]	-	0.0003	-	0.7987	-	0.0167	-	0.0268

[1] All statistics, including the difference in LSM (least square means), were from ANCOVA model for the comparison between each belimumab dose and placebo, adjusted for the baseline PGA score, baseline SELENA-SLEDAI score (≤ 9 vs ≥ 10), baseline proteinuria level (< 2 g/24 hour vs ≥ 2 g/24 hour equivalent) and race (African descent or indigenous-American descent vs other). For pooled data analysis, study was also included as an additional covariate.

The mean change in PGA at week 52 was submitted as an additional secondary outcome. The results are shown in Table 17.

Table 17: Mean change in PGA score at week 52 (taken from MS Table 5.18)

Other secondary endpoints Week 52	BLISS-52		BLISS-76		Pooled Total Population		High Disease Activity Subgroup	
	Placebo N = 287	10mg/kg N = 290	Placebo N = 275	10mg/kg N = 273	Placebo N = 562	10mg/kg N = 563	Placebo N = 203	10mg/kg N = 193
Mean ± SE	-0.48 ± 0.04	-0.67 ± 0.04	-0.46 ± 0.04	-0.49 ± 0.04	-0.47 ± 0.03	-0.58 ± 0.03	-0.41 ± 0.05	-0.62 ± 0.05
LS Mean ± SE[1]	-0.38 ± 0.05	-0.57 ± 0.05	-0.47 ± 0.06	-0.55 ± 0.06	-0.40 ± 0.04	-0.54 ± 0.04	-0.36 ± 0.06	-0.59 ± 0.06
P-value[1]	-	0.0001	-	0.1159	-	< 0.0001	-	0.0003

[1] All statistics, including the difference in LSM (least square means), were from ANCOVA model for the comparison between each belimumab dose and placebo, adjusted for the baseline PGA score, baseline SELENA-SLEDAI score (≤ 9 vs ≥ 10), baseline proteinuria level (< 2 g/24 hour vs ≥ 2 g/24 hour equivalent) and race (African descent indigenous-American descent vs other). For pooled data analysis, study was also included as an additional covariate.

In BLISS-52 a larger improvement in PGA score was observed for the 10mg/kg group than for placebo (P = 0.0001) whereas in BLISS-76 the difference between treatments was trivial (P = 0.115). For the pooled populations 10mg/kg was superior to placebo (P = 0.0003).

The HGS Briefing Document to the FDA[5] provided graphed results for mean change in PGA through successive clinic visits for all three randomised groups. These are shown below in Figure 5 for BLISS-76.

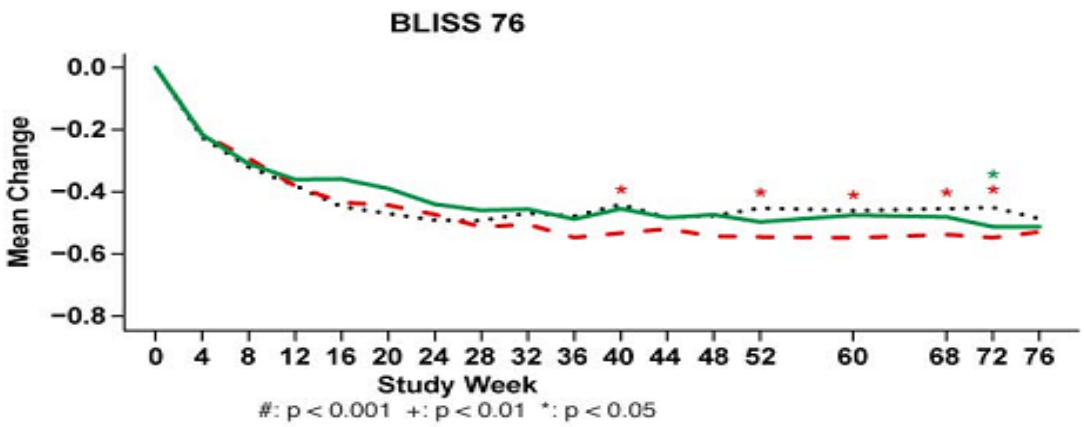


Figure 5: Mean change in PGA score in BLISS-76

Across 76 weeks of follow up in BLISS-76 the 1mg/kg dose regimen appeared to outperform the 10mg/kg regimen. Baseline differences (MS Table 5.9) were similar between treatment groups.

The mean change in SF-36 PCS scores at week 24 relative to baseline, a major secondary outcome, showed little difference between belimumab and placebo groups in BLISS-52 (P = 0.8870), or in BLISS-76 (P = 0.6601), or in the Target population pooled across trials (P = 0.4276).

Change in SF-36 PCS scores at week 52 was specified as a non-major secondary outcome. No significant improvement was observed for BLISS-76 or Target populations (P = 0.5134 and P = 0.1124, respectively) however in BLISS-52 the difference between belimumab and placebo arms (4.18 vs. 2.96) was sufficient to reach statistical significance (P = 0.0247).

Reduction in steroid use between weeks 40 and 52 for those patients receiving ≥ 7.5 mg/day prednisone at baseline was specified as a major secondary outcome. The results submitted summarised in Table 18.

Table 18: Prednisone reduction Weeks 40 through 52 – Phase 3 trials

	BLISS-52		BLISS-76		Pooled Total Population[4]		High Disease Activity Subgroup Pooled Total		High Disease Activity Subgroup BLISS-52		High Disease Activity Subgroup BLISS-76	
	Placebo N = 192	**10 mg/kg N = 204**	**Placebo N = 126**	**10 mg/kg N = 120**	**Placebo N = 318**	**10 mg/kg N = 324**	**Placebo N = 126**	**10 mg/kg N = 126**	**Placebo N = 76**	**10 mg/kg N = 81**	**Placebo N = 50**	**10 mg/kg N = 45**
No. %[1] Response[2]	23 12.0%	38 18.6%	16 12.7%	20 16.7%	39 12.3%	58 17.9%	9 7.1%	20 15.9%	4 5.3%	15 18.5%	5 10.0%	5 11.1%
Observed difference vs Placebo	-	6.65	-	3.97	-	5.64	-	8.73	-	13.5	-	1.1
OR (95% CI)[3] vs placebo	-	1.75 (0.99, 3.08)	-	1.26 (0.61, 2.60)	-	1.57 (1.01, 2.45)	-	2.43 (1.05, 5.65)	-	4.11 (1.29, 13.2)	-	0.88 (0.21, 3.60)
P-value[3]	-	0.0526	-	0.5323	-	0.0451	-	0.0389	-	0.0171	-	0.8586

[1] Includes only subjects with baseline prednisone > 7.5 mg/day
[2] Any subject who withdrew from the study prior to the Day364 (Week 52) visit, missed the Day 364 (Week 52) visit (± 28 day window allowed), and/or received a protocol-prohibited medication or a dose of allowable (but protocol-restricted) medication that resulted in treatment failure designation prior to the Day 364 (Week 52) visit was considered a treatment failure for prednisone reduction
[3] Odds Ratio (95% confidence interval) and p-value were from logistic regression for the comparison between each belimumab dose and placebo with covariates. For individual studies, the covariates include baseline prednisone level, baseline SELENA-SLEDAI score (≤ 9 vs ≥ 10), baseline proteinuria level (< 2 g/24 hour vs ≥ 2 g/24 hour equivalent) and race (African descent or indigenous-American descent vs other). For pooled data analysis, study was also included as an additional covariate
[4] Obtained from a logistic regression by adding study and the treatment-by-study interaction to the above model

In BLISS-52 and BLISS-76 at baseline 68.6% and 44.9% of patients respectively were receiving ≥ 7.5 mg/day prednisone. The percentage that reduced steroid use in weeks 40 to 52 by the pre-specified amount was greater in the belimumab arm than the placebo arm in both trials, however the difference (belimumab vs. placebo) failed to reach statistical significance in either trial: 18.6% vs. 12.0% in BLISS-52 (P = 0.0526 from logistic regression including baseline covariates) and 16.7% vs. 12.7% in BLISS-76 (P = 0.5323).

For the Target or HDAP pooled across trials 15.9% and 7.1% reduced steroid use in the 10mg/kg belimumab and placebo groups respectively (P = 0.0389 from logistic regression). The results from BLISS-52 supported belimumab (P = 0.171) whereas in BLISS-76 differences between treatments were trivial (P = 0.8586). The HGS Briefing Document to the FDA[5] provided results for reduction in steroid use for all three treatment arms. Table 9-16 from the HGS Briefing Document is shown in Figure 6.

Table 9-16 Reduction in steroid use - Phase 3 trials[1]

	BLISS 52			BLISS 76			Both Studies		
	Placebo N = 192	1 mg/kg N = 204	10 mg/kg N = 204	Placebo N = 126	1 mg/kg N = 130	10 mg/kg N = 120	Placebo N = 318	1 mg/kg N = 334	10 mg/kg N = 324
No. (%) Response[2]	23 (12.0%)	42 (20.6%)	38 (18.6%)	16 (12.7%)	25 (19.2%)	20 (16.7%)	39 (12.3%)	67 (20.1%)	58 (17.9%)
OR (95% CI) vs placebo		1.89 (1.08, 3.31)	1.75 (0.99, 3.08)		1.57 (0.78, 3.14)	1.26 (0.61, 2.60)		1.77 (1.15, 2.73)	1.57 (1.01, 2.45)
P-value[3]		0.0252	0.0526		0.2034	0.5323		0.0097	0.0451
Treatment by study interaction p-value[4]		NA	NA		NA	NA		0.7020	0.5177

1 Average prednisone (equivalent) dose reduced by ≥ 25% from baseline to ≤ 7.5 mg/day during Weeks 40-52 in the subgroup of patients who were receiving > 7.5 mg/day of prednisone at baseline

2 Any patient who withdrew from the study prior to the Day 364 (Week 52) visit, missed the Day 364 (Week 52) visit (± 28 day window allowed), and/or received a protocol-prohibited medication or a dose of allowable (but protocol-restricted) medication that resulted in treatment failure designation prior to the Day 364 (Week 52) visit was considered a treatment failure for prednisone reduction.

3 Odds ratio (95% confidence interval) and p-values were from logistic regression for the comparison between each belimumab dose and placebo with covariates. For individual studies, the covariates include baseline prednisone level, baseline SELENA SLEDAI score (≤ 9 vs ≥ 10), baseline proteinuria level (< 2 g/24 hour vs ≥ 2 g/24 hour equivalent) and race (African descent or indigenous-American descent vs other). For pooled data analysis, study was also included as an additional covariate. P-values nominal.

4 Obtained from a logistic regression by adding study and the treatment-by-study interaction to the above model.

Figure 6: Reduction in steroid use Phase II trials (Taken from HGS Table 9-16)

It is noticeable that again there was little difference in effectiveness between the 1mg/kg and 10mg/kg dose regimens, in BLISS-76 a better outcome was recorded with 1mg/g than with 10mg/kg, and that the results from BLISS-52 were more strongly supportive of belimumab than those from BLISS-76.

Further secondary outcomes submitted

Flares

Time to first flare and to first severe flare was reported in MS Figures 5.9 to 5.13.

In BLISS-52 the time to first flare was delayed by 10 mg/kg belimumab relative to placebo (HR 0.76, 95% CI: 0.63 – 0.91, P = 0.0036) (Figure 7).

Figure 7: Time to first flare; BLISS-52 (Taken from MS Figure 5.10)

In BLISS-76 there was no difference between groups in time to first flare (P = 0.4796; Figure 8).

Figure 8: Time to first flare; BLISS-76 (Taken from MS Figure 5.11)

When the whole populations from the BLISS trials were pooled the difference between treatments reached statistical significance (P = 0.0120; Figure 9).

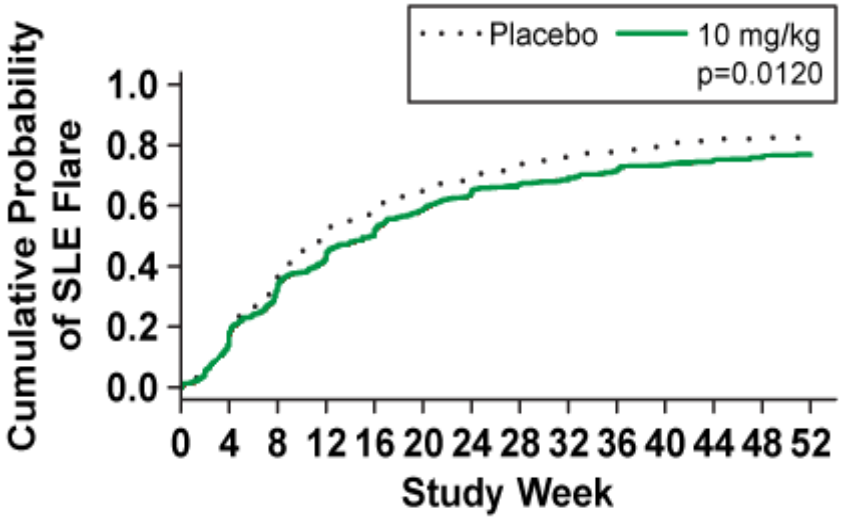


Figure 9: Time to first flare; pooled whole populations (Taken from MS Figure 5.12)

For the high disease activity Target population pooled across trials, belimumab significantly delayed time to first flare relative to placebo (P = 0.007; Figure 10).

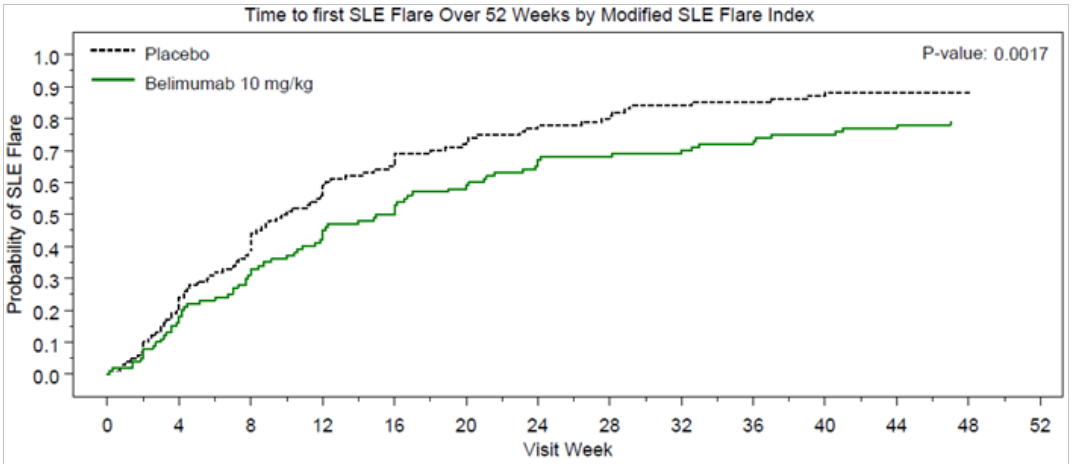


Figure 10: Time to first flare; pooled Target populations (Taken from MS Figure 5.13)

In BLISS-52 the time to first severe flare was delayed by 10 mg/kg belimumab relative to placebo P = 0.0055; Figure 11).

Figure 11: Time to first severe flare; BLISS-52 (Taken from MS Figure 5.10)

In BLISS-76 belimumab somewhat delayed time to first severe flare in (HR 0.72, 95% CI 0.50–1.05, P = 0.0867; Figure 12).

Figure 12: Time to first severe flare; BLISS-76 (Taken from MS Figure 5.11)

When the whole populations from the BLISS trials were pooled the difference between treatments for time to first severe reached statistical significance (P = 0.0011; Figure 13).

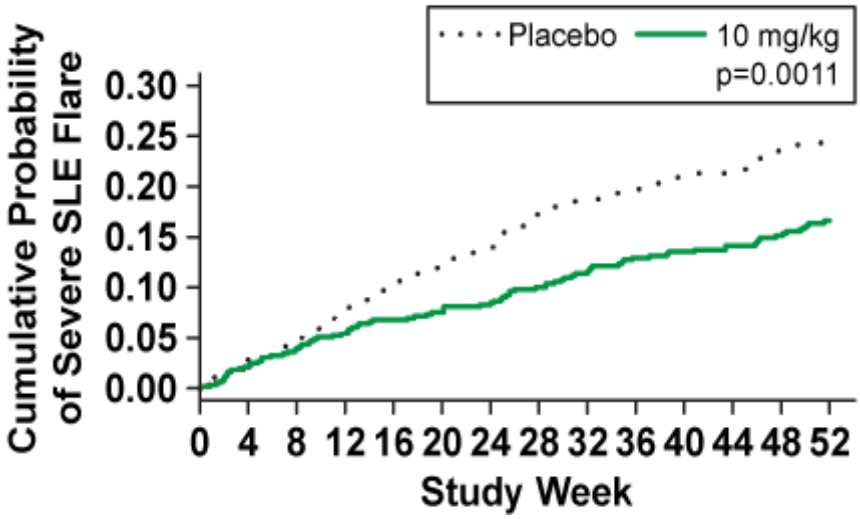


Figure 13: Time to first severe flare; pooled whole population (Taken from MS Figure 5.12)

For the high disease activity Target population pooled across trials, belimumab significantly delayed time to first severe flare relative to placebo (P = 0.0028; Figure 14).

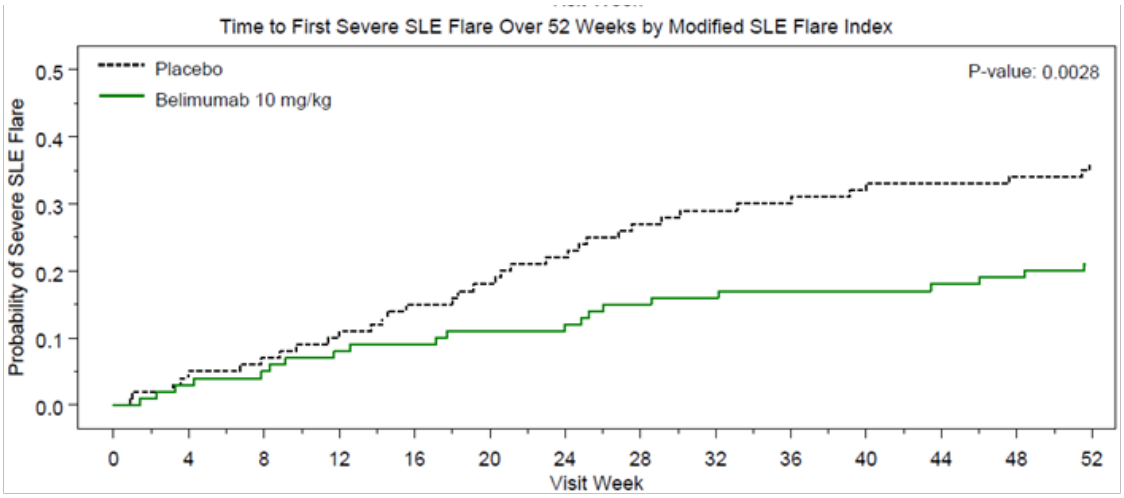


Figure 14: Time to first severe flare; pooled Target population (Taken from MS Figure 5.13)

The HGS Briefing Document to the FDA[5] provided the graphs shown in Figure 15 depicting results for all three treatment arms.

Figure 15: Time to first flare (taken from HGS Briefing Document to the FDA)

It is noticeable that in both trials the 1 mg/kg belimumab dose regimen was as effective as the 10 mg/kg dose regimen in extending time to first flare, and that for BLISS-76 this also applies for severe flares. For both flares and severe flares the results from BLISS-52 were more supportive of belimumab than those from BLISS-76.

SLICC/ACR Damage Index

There was no meaningful difference between the belimumab and placebo groups in the change in SLICC/ACR Damage Index at Week 52 compared with baseline.

FACIT-fatigue index

The mean change FACIT fatigue score from baseline was reported in MS Figures 5.14 to 5.17 shown in Figure 16.

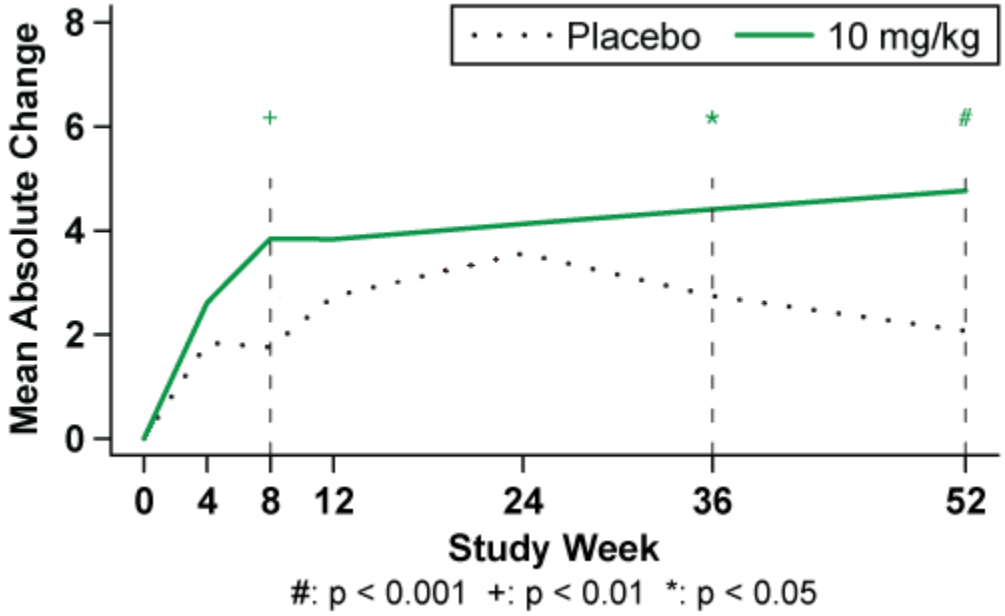


Figure 16: Mean change in FACIT-Fatigue score – BLISS-52 (Taken from MS Figure 5.14)

At week 52 relative to baseline the belimumab group had greater improvement in FACIT-Fatigue score than the placebo group (4.8 belimumab and 2.1 placebo in BLISS 52; 4.6 and 2.9 in BLISS-76). The difference was significant for BLISS-52 ($P < 0.001$) but not for BLISS-76 ($P \geq 0.05$) (Figure 17 and Figure 18).

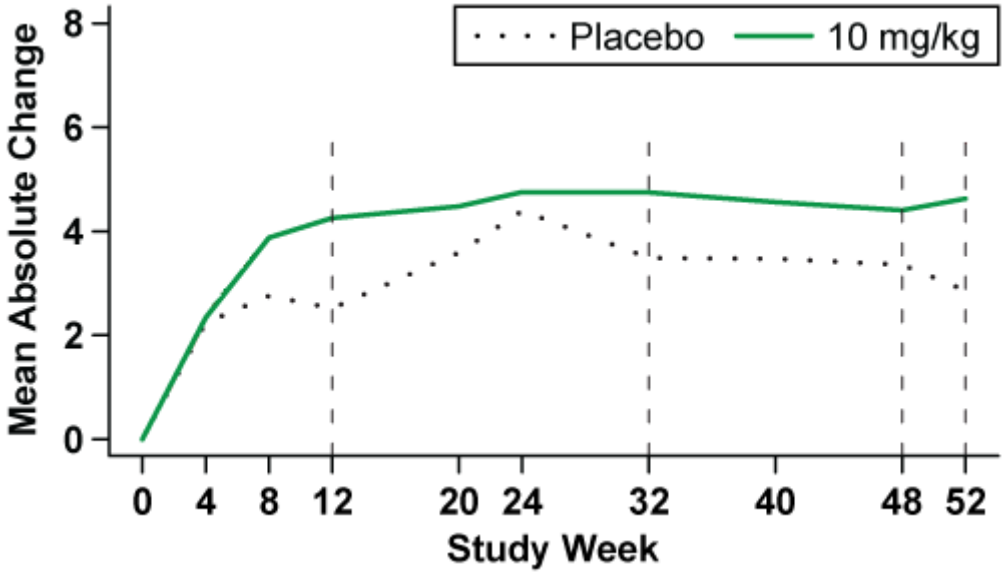


Figure 17: Mean change in FACIT-Fatigue score – BLISS-76 (Taken from MS Figure 5.15)

For the whole population pooled across trials the difference was statistically in favour of belimumab at week 52 (Figure 18).

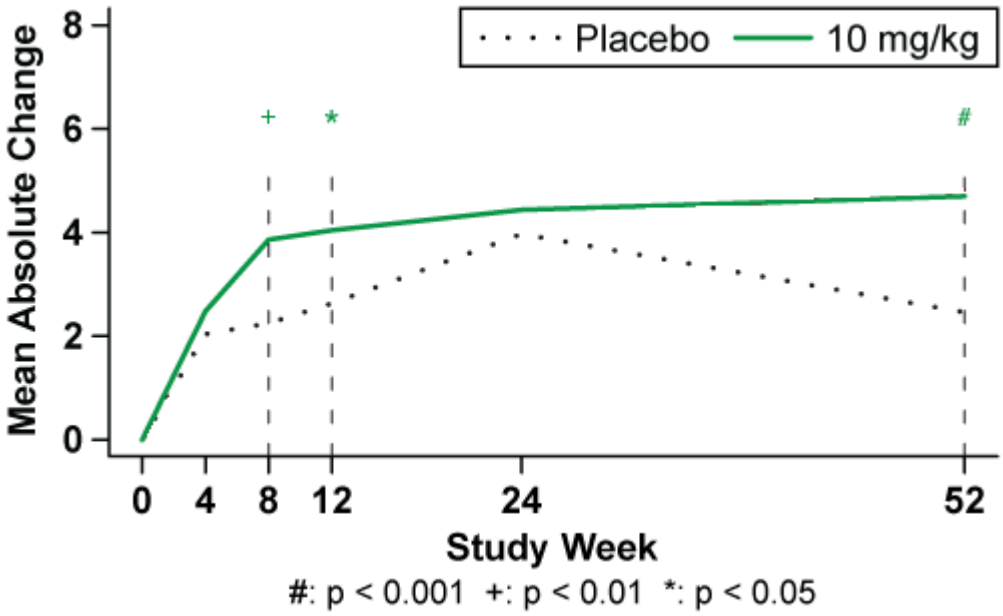


Figure 18: Mean change in FACIT-Fatigue – Pooled Total Population (Taken from MS Figure 5.16)

While for the target population pooled across trials at weeks 8 and 12 the difference between groups was statistically in support of belimumab (P < 0.05) however the difference between groups then diminished; at week 52 there was no longer a significant difference (see Figure 19).

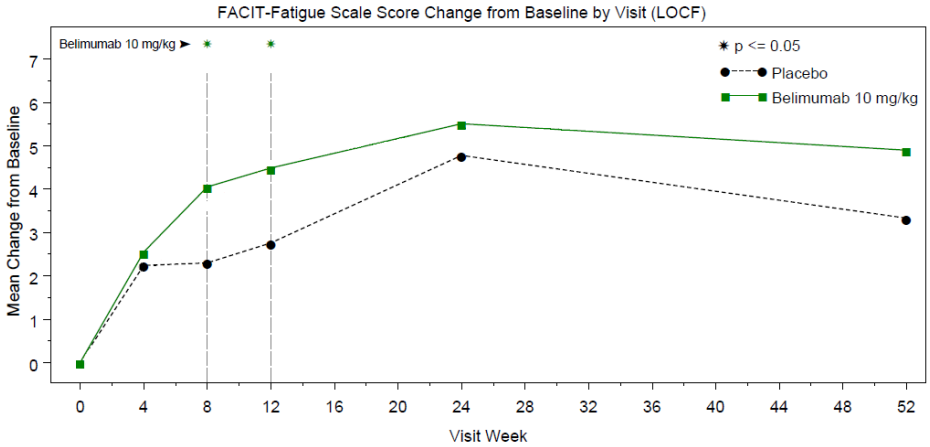


Figure 19: Mean change in FACIT-Fatigue – pooled Target population (Taken from MS Figure 5.17)

The HGS Briefing Document to the FDA provided results for all three treatment arms at week 52. In BLISS-52 the 10mg/kg dose was more effective than the 1 mg/kg but for BLISS-76

the reverse was the case. The BLISS-76 result is shown in Figure 20 together the mean change in SF-36 Vitality domain score.

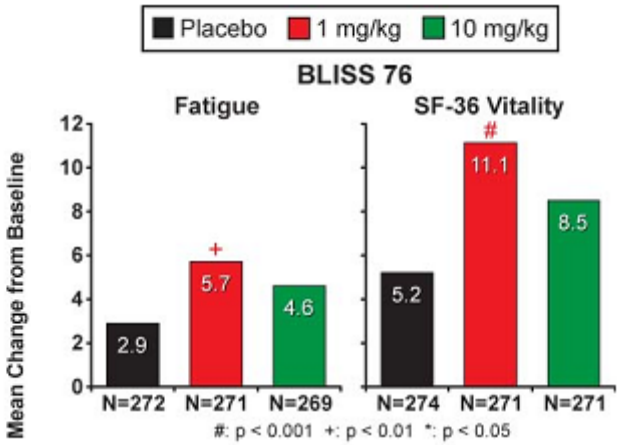


Figure 20: Mean change in FACIT and SF-36 vitality score by week 52 (Taken from HGS Briefing Document to FDA see Figure 9-35)

EQ-5D

There was no significant difference between belimumab and placebo in the absolute change of EQ-5D score from baseline in either trial or pooled total populations during clinic visits. The results for the 10 mg/kg belimumab and placebo groups in BLISS-76 were indistinguishable. For the pooled target population the difference between 10 mg/kg and placebo groups reached statistical significance in favour of belimumab at week 24 ($P \leq 0.01$), but the difference had almost completely faded by week 52 MS Figure 5.21 Page 135).

Results for the mean change in SELENA SLEDAI score from baseline at week 52 were submitted in MS Table 5.17 (Page 113) and clarification response Table A6.1 and are summarised in Table 19. There was no significant difference between belimumab and placebo in the absolute change of EQ-5D score from baseline in either trial or pooled total populations during clinic visits. The results for the 10mg/kg belimumab and placebo groups in BLISS-76 were indistinguishable. For the pooled Target population the difference between 10mg/kg and placebo groups reached statistical significance in favour of belimumab at week 24 ($P \leq 0.01$), but the difference had almost completely faded by week 52 MS Figure 5.21 Page 135).

Results for the mean change in SELENA SLEDAI score from baseline at week 52 were submitted in MS Table 5.17 (Page 113) and clarification response Table A6.1 and are summarised in Table 19.

Table 19: Mean change and mean percent change in SLEDAI score week 52

	BLISS-52		BLISS-76		Pooled Total Population[2]		High Disease Activity Subgroup Pooled Total		High Disease Activity Subgroup BLISS-52		High Disease Activity Subgroup BLISS-76	
Change from Baseline at Week 52	**Placebo N = 287**	**10mg/kg N = 290**	**Placebo N = 275**	**10mg/kg N = 273**	**Placebo N = 562**	**10mg/kg N = 563**	**Placebo N = 203**	**10mg/kg N = 193**	**Placebo N = 107**	**10 mg/kg N = 112**	**Placebo N = 96**	**10 mg/kg N = 81**
Mean change from baseline (± SE)	-3.57 ± 0.24	4.97 ± 0.27	-2.77 ± 0.25	-3.70 ± 0.27	-3.18 ± 0.18	-4.36 ± 0.19	-4.1 ± 0.3	-5.8 ± 0.3	-4.1 ± 0.4	-6.3 ± 0.5	-4.0 ± 0.5	-5.2 ± 0.5
P-value[1]	-	< 0.0001	-	0.0063	-	< 0.0001	-	0.0005	-	0.0008	-	0.1705
Mean % change (± SE)	-34.76 ± 2.50	-45.60 ± 2.45	-25.97 ± 2.72	-35.94 ± 2.80	-30.47 ± 1.85	-40.93 ± 1.86	-30.5 (2.3)	-45.5 (2.4)	-	-	-	-
P-value[1]	-	0.0018	-	0.0073	-	< 0.0001	-	< 0.0001	-	-	-	-

[1] ANCOVA model for the comparison between each belimumab dose and placebo, adjusted for baseline SELENA SLEDAI score (≤ 9 vs. ≥ 10), baseline proteinuria level (< 2 g/24 hour vs. ≥ 2 g/24 hour equivalent) and race (African descent or indigenous-American descent vs. other). For pooled data analysis, study was also included as an additional covariate

[2] No treatment-by-study interactions observed (all p-values > 0.367)

Both absolute SLEDAI score reduction from baseline, and percent reduction relative to baseline score, were greater for the 10mg/kg group than for the placebo group; this was consistent and significant for the whole BLISS population (separately by trial and for pooled populations) and for the pooled Target or high disease activity population. For the whole population, results favoured belimumab more strongly in BLISS-52 than BLISS-76. The by-trial results for the Target population are shown below. They indicate stronger support for belimumab in BLISS-52 in which the difference between groups in absolute reduction in SLEDAI score was about double that in BLISS-76 in which the difference between groups was not significant (P = 0.1705).

The HGS Briefing Document to the FDA[5] (see Figure 21) showed the percentage change in SLEDAI score (relative to baseline) throughout the two trials; this is reproduced in Figure 21 for the mean change in FACIT and SF-36 vitality score by week 52.

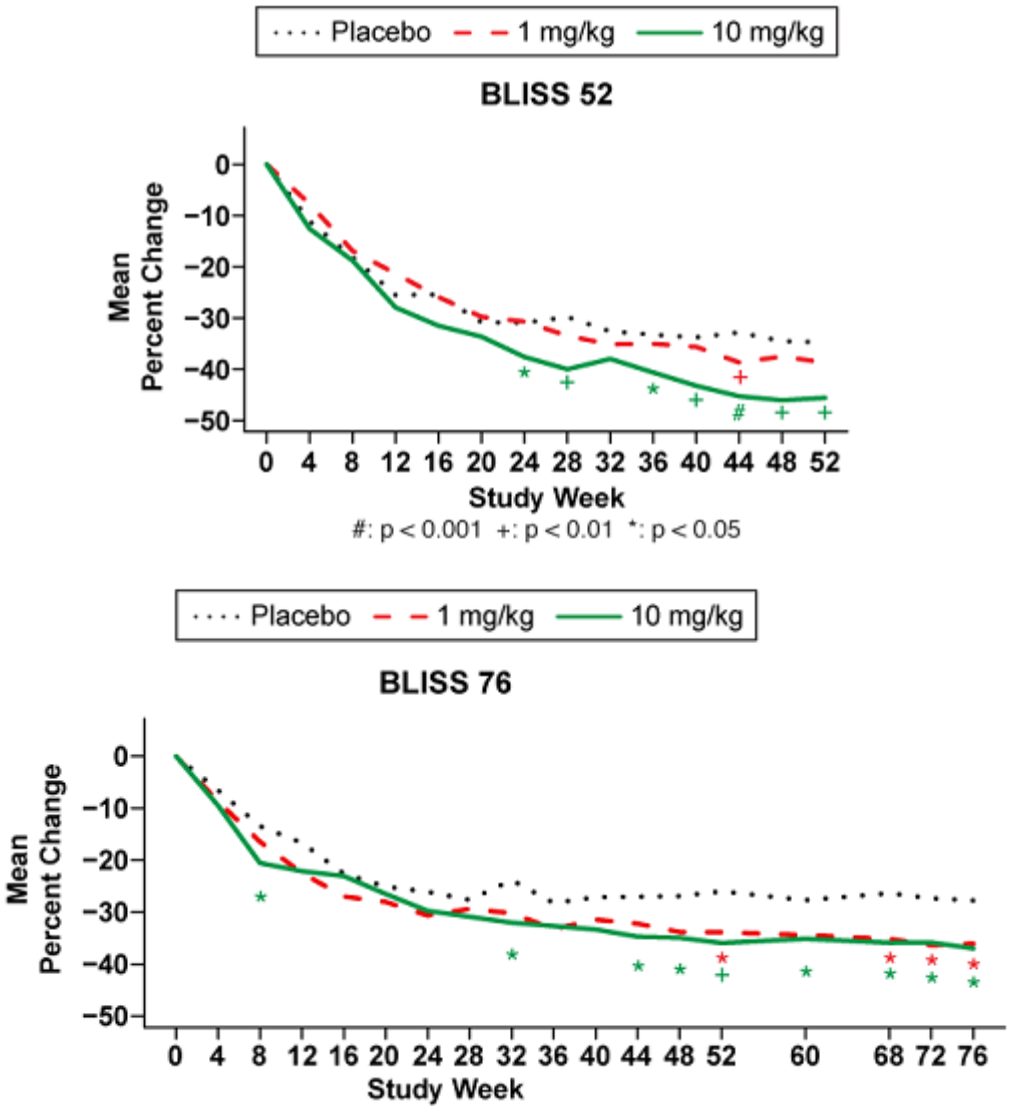


Figure 21: Percentage change in SLEDAI score according to treatment arm (Taken from HGS Briefing Document Figure 9-29)

These results support a dose response relationship in BLISS-52, but the difference between 1mg/kg and 10mg/kg dose regimens in the BLISS-76 trial is relatively trivial.

Safety

The submission pooled results from three RCTs: BLISS-52, BLISS-76 and LBSL02. LBSL02 lasted 52 weeks, preceded the BLISS trials, and was conducted in North America (98% patients from the USA), and did not employ the SRI composite outcome measure. The LBSL02 trial randomised 449 patients to one of four treatments: SoC + placebo, SoC + 1mg/kg belimumab, SoC + 4mg/kg belimumab, and SoC + 10mg/kg belimumab. Although

all patients had a history of auto-immunity, at recruitment 30% currently lacked anti-nuclear antibodies.

There were 15 deaths during the controlled phase of the three trials; 3 in the placebo group (n=675), and 12 in the belimumab groups (n=1458) with 6 each in the 10mg/kg and 1mg/kg groups respectively. One death in the 1mg/kg belimumab group followed 15 weeks after the patient stopped belimumab treatment. The causes of death were various and are listed in Table 20 (based on FDA Briefing Package, Table 34).[3] When deaths are rated according to exposure these results translate to: 0.43/100 patient years for placebo (95% CI: 0.09, 1.27) and 0.79/100 patient years for belimumab (95% CI: 0.41, 1.38).[3] There were two completed suicides in the belimumab groups (none in placebo); a further suicide was observed during the LBSL99 extension study. These were not judged to be associated with belimumab since the patients concerned had a history of depression and SLE is associated with an increased risk of depression and suicide.

Table 20: Deaths occurring during controlled phase of belimumab RCTs

Study group	Age/Sex	Cause	Days from 1'st infusion	Days from last infusion	Pertinent History
Placebo	45yo/F	Myocardial Infarction	328	19	Presented to ER with new onset chest and epigastric pain and had a cardiopulmonary arrest.
Placebo	25yo/F	Cardiac Arrest, secondary to sepsis	70	11	Concomitant Meds: Prednisolone, methotrexate, diclofenac and ibuprofen. Developed bacterial gastroenteritis and dehydration complicated by vasculitis and became septic (blood culture positive for Staph. Saprophyticus) despite antibiotics and supportive medical care.
Placebo	18yo/F	Unknown	225	84	Hospitalized 2 months prior to death for acute abdominal pain secondary to portal/mesenteric/renal vein and vena cava thrombosis and acute pancreatitis.
Belimumab 1mg/kg	43yo/F	Suicide	32	20	H/O Depression on antidepressant (citalopram). Reported to have worsening depression prior to committing suicide.
Belimumab 1mg/kg	46yo/F	Unknown	56	28	H/O Asthma, clostridial gastroenteritis, eosinophilia and QT prolongation on EKG. Concomitant Meds: ibuprofen, hydroxychloroquine, mycophenolate, prednisone and lisinopril. Pt. developed nausea, vomiting and weakness while camping and was found to be dehydrated due to unspecified gastrointestinal illness at local ER where she died despite resuscitative measures.
Belimumab 1mg/kg	52yo/F	Ovarian cancer	21	7	Positive family H/O ovarian cancer. H/O Vaginal bleeding prior to study entry that evolved to include left lower abdominal pain, vaginal pain, pelvic cramping and diarrhea by the 9th study medication that was followed by a diagnosis of advanced ovarian cancer on laporotomy.
Belimumab 1mg/kg	32yo/F	Sepsis, secondary to cellulitis	13	13	Concomitant Meds: Methylprednisolone, mycophenolate, thalidomide, and ibuprofen. Developed cellulitis and died as a result of sepsis despite antibiotics and supportive medical care.
Belimumab 1mg/kg	58yo/F	Ischemic stroke	345	34	H/O hypertension. Anti-cardiolipin antibody negative at screening. Concomitant meds: Prednisolone, hydroxychloroquine, bioprolol.
Belimumab 1mg/kg	25yo/F	Respiratory failure /SLE flare	216	104	Patient died due to respiratory arrest more than 15 weeks after the patient discontinued the trial due to acute renal failure. Post study withdrawal, the patient was hospitalized and experienced oliguria, uremic syndrome, sepsis, polyserositis, ascites, intestinal edema, anemia, and alveolar hemorrhage.
Belimumab 10mg/kg	40yo/F	Respiratory failure secondary to sepsis	257	33	Pt. developed aspiration pneumonia status post seizure, became septic and died due to respiratory failure despite antibiotics and aggressive supportive medical care (respirator).
Belimumab 10mg/kg	47yo/F	Cardiac arrest (SLE flare)	77	21	H/O Diabetes mellitus, pericardial excision, serositis, antiphospholipid syndrome, pulmonary hypertension, and heart failure. Concomitant Meds: Azathioprine, methotrexate and prednisone. Hospitalized after c/o severe headache with vomiting associated with fever, chills and productive cough with bilateral pleural effusions and lymphopenia attributed to SLE flare with CNS involvement. She was treated with corticosteroids and NSAIDs but died due to cardiac arrest.
Belimumab 10mg/kg	53yo/F	Bacterial sepsis	331	25	H/O Obesity, pulmonary fibrosis. Developed septic shock (blood cultures positive for MRSA) and multi-organ failure secondary to infected herpes zoster lesions despite antibiotics. Concomitant meds: Methyplprednisone, azathioprine, chloroquine, salbutamol, acenocoumarol, sertraline, and omeprazole.
Belimumab 10mg/kg	20yo/F	Infectious diarrhea	336	28	Had SLE flare with cutaneous vasculitis and hypochromic anemia. Started on antibiotics and increased corticosteroids but developed infectious diarrhea and died en route to hospital. Concomitant meds: Prednisolone, azathioprine, hydroxychloroquine, levofloxacin, iron, ciprofloxin/tinidazole, and fluconazole.
Belimumab 10mg/kg	23yo/F	Suicide	272	13	H/O Depressed mood and psychotic disorder; autoimmune thyroiditis, and drug-induced hepatitis. Committed suicide following conflict with parent. Concomitant meds: methylprednisone, azathioprine, hydroxychloroquine, meloxicam, levothyroxine, and rebamipide.
Belimumab 10mg/kg	33yo/F	Respiratory Failure From Presumed Pulmonary Embolus	128	8	H/O chronic cholecystitis. Pt. developed dyspnea eight days after her last study infusion and died en route to the hospital. (No autopsy.) Concomitant meds: Prednisone, levothyroxine, and ceftriaxone.

Adverse events

In all treatment groups > 90% of patients experienced at least 1 AE. The most commonly occurring AEs were headache, upper respiratory tract infection, arthralgia, nausea, UTI, diarrhoea and fatigue.

The percentage of patients experiencing at least one serious AE and at least one serious AE was very similar between placebo and belimumab groups ranging 13.5% to 18.6%, there was a very slight numerical excess with belimumab. The most frequent serious AEs (≥ 1% in any treatment group) were pneumonia, pyrexia, UTI, cholelithiasis, and cellulitis. The percentage of patients experiencing at least one severe AE was 15.4% for the placebo group and 16% across the belimumab groups; the most common severe adverse events were not identified.

Infections

Infections occurred slightly more frequently in patients treated with belimumab compared to placebo. The most frequent infections were URTI, UTI, nasopharyngitis, sinusitis, and bronchitis.

Infusion / hypersensitivity reactions

Occurrence of infusion plus hypersensitivity reactions was similar between belimumab and placebo-treated patients (17% and 14.7%, respectively). Of 1458 belimumab treated patients, 15 experienced hypersensitivity reactions on the day of infusion compared to one of 675 placebo-treated patients.[3] Five discontinuations resulted from hypersensitivity reactions amongst 1458 belimumab patients and none among 675 patients receiving placebo.

The most frequent infections were URTI, UTI, nasopharyngitis, sinusitis, and bronchitis. Of these, nasopharyngitis and bronchitis occurred more commonly with belimumab treatment compared to placebo. Two opportunistic infections occurred, both in the belimumab 10mg/kg group: disseminated CMV infection on day 62; and an Acinetobacter bacteremia on day 15. Four infections were related to deaths: sepsis (placebo group); infectious diarrhea (belimumab 10mg/kg group); cutaneous infection leading to sepsis (belimumab 10mg/kg group); and cellulitis leading to sepsis (belimumab 1mg/kg group).[3]

4.2.6 Pooling of trial data

NICE requests that: "*For each outcome for each included RCT, the following information should be provided*" …."*The size of the effect*" … "*The unit of measurement*".

The submission pooled results from two trials, both for the whole BLISS populations and for the Target populations. The manufacturer considered that the pooled trial results were most

appropriate for the decision problem, and importantly it was pooled results for both populations that were entered into the economic model. The MS stated as shown in Box 11.

Box 11: Taken from Page 21 and 97 of the MS

"For the purpose of this submission pooled efficacy from the two pivotal Phase III studies is considered most relevant to the decision problem." Refer to MS Page 21 AND "Pooling is appropriate given that the trials were essentially identical in design and in the analysis of the primary endpoint, the p-values for the treatment-by-study interaction were not significant (interaction p-values > 0.5)." Refer to MS Page 97

Comment

The trials were conducted according to very similar protocols and used the same primary end point so the lack of significant treatment-by-study interaction was not surprising; see section 4.2.4 of this report.

The submission initially supplied only pooled results for the Target population, therefore the ERG requested "by-trial" results and further justification for pooling. The response to this request is shown in Box 12 (for the full response see Appendix 5).

Box 12: From the manufacturer's clarification response

"….one must then determine whether the relative treatment effect is different in one study compared with the other study when evaluating whether two studies are similar enough to pool. Each of the Phase 3 studies achieved statistical significance for belimumab 10mg/kg on the pre-specified primary endpoint of SRI response at Week 52; therefore, these nearly identical, studies provide independent replication of results."

Comment

There remain doubts as to whether the pooled trial results are relevant for patients in England and Wales. The BLISS trials were run globally, recruited 1684 patients in 226 centres across 32 countries and involved a large number of different investigators. Although both trials achieved statistical significance on the primary end point, they were dissimilar in underlying patient groups e.g. by ethnicity and in effect size for almost all outcomes, with BLISS-52

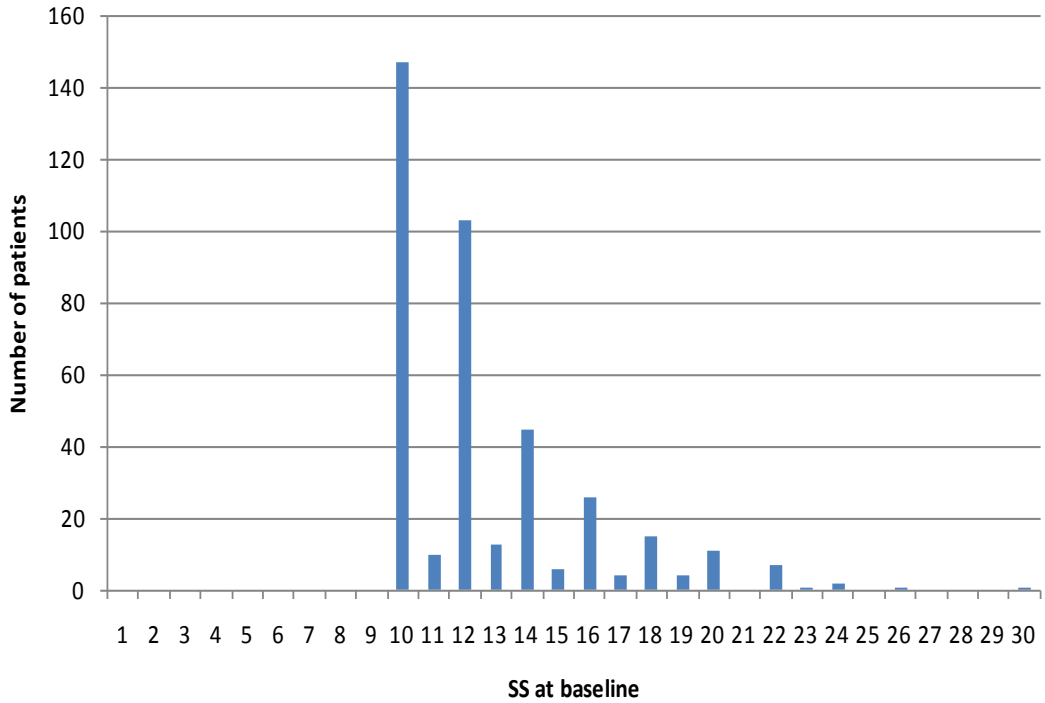
providing stronger results than BLISS-76; furthermore there were inconsistencies with regard to an expected dose response relationship.

The ERG is concerned that the pooled results are mainly driven by those from the BLISS-52 trial (conducted in Pacific-Asia and South America) while results in the BLISS-76 trial, conducted in North America and Europe, were only marginally in favour of belimumab relative to placebo and reached statistical significance only for two overlapping outcomes (SRI responders at week 52, and percentage of patients with ≥ 4point reduction in SLEDAI score at week 52 which itself is a component of the SRI). The extent to which these concerns extend to the target population was not possible to gauge from the initial submission because only pooled results were presented. The ERG therefore requested clarification on trial specific target populations and the manufacturer's justification for pooling across trials.

For the target population there was again a greater contribution from BLISS-52 to the pooled results both in terms of number of patients (BLISS-52 contributed 55% of whole target population and 58% of those that received belimumab) and in effectiveness (BLISS-52 provided greater effect sizes compared to BLISS-76 for SRI week 52, modified SRI week 52, percentage with SLEDAI reduction by ≥ 4points, SLEDAI mean change by week 52, no new BILAG 1A/2B, no worsening in PGA and reduction in steroid usage weeks 40 to 52). The ERG therefore remain concerned that the pooled trial results dilute the rather less positive findings most relevant to the decision problem by including additional data from a less relevant population, and that target population results by trial should have been included in sensitivity analysis in the economic model.

Trial baseline SLEDAI scores used in economic model

The manufacturer's economic analysis (section 5) made use of data from an SLE cohort studied at JHU so as to model cost effectiveness of belimumab for the whole and Target populations from the BLISS trials. The JHU cohort experienced relatively mild disease compared to patients in BLISS and particularly compared to the BLISS the high disease activity Target population. During the clarification process the ERG requested the distribution of SLEDAI scores at year one and last year of observation for patients in the JHU cohort. The SLEDAI scores shown in Figure 22 illustrates the differences between Target and JHU populations, (year one and last year scores are shown for JHU cohort, Figure B17.2 of the clarification document, the clarification response did not make clear which was year one and which last year).



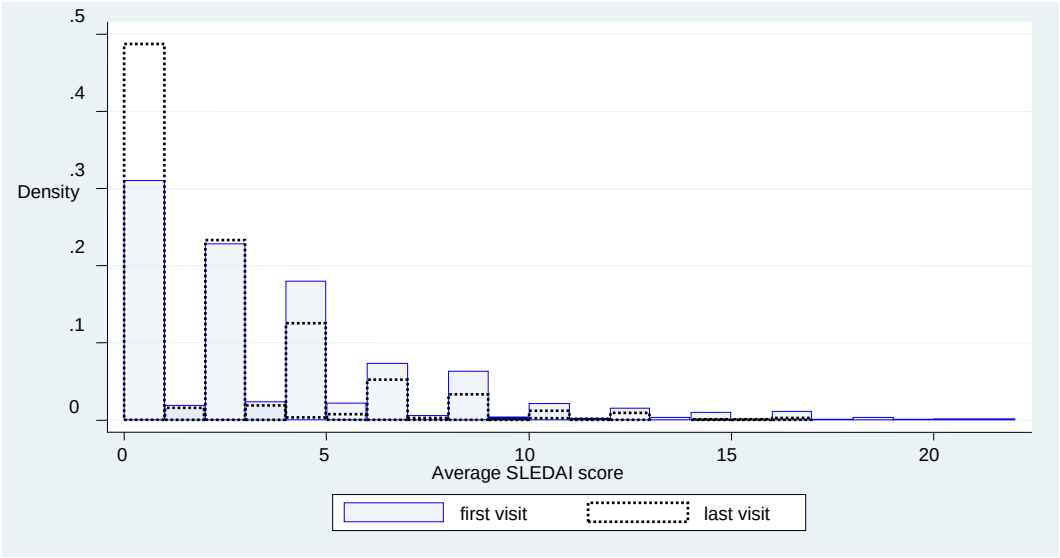


Figure 22: SLEDAI scores for Target and JHU populations (from clarification document)

The median follow up for the JHU cohort was 6.9 years. During this time organ damage progression was monitored and for economic analysis this was modelled (using parametric time event analyses) so as to be able to predict organ damage progression for BLISS patients according to their observed SLEDAI changes. Yet for the JHU cohort the difference between year one and last year in average SLEDAI scores is small. The ERG considers that this indicates some inadequacy in using the short term measure of disease activity (i.e. SLEDAI) to model how a group of patients will progress to organ damage.

4.2.7 Conclusions

Efficacy evidence came from two multicentre international industry sponsored RCTs (BLISS-52 and BLISS-76) comparing SoC plus belimumab with SoC plus placebo; each trial had three arms: placebo, 1mg/kg belimumab and 10mg/kg belimumab dose regimens. Data for the 1mg/kg arms was excluded from the submission, but results available in the public domain were considered in the ERG's assessment. Outcomes for six populations were presented: whole populations from BLISS trials, whole populations pooled across BLISS trials, Target population from BLISS trials and Target populations pooled across BLISS trials.

The Target population was a high disease activity subgroup identified from post hoc exploration of effectiveness of the primary outcome. There were more noticeable within-trial baseline imbalances (10mg/kg vs. placebo) for the Target population than for the whole population. The Target population results are not necessarily equivalent to those that would be obtained from a randomised trial in this population.

The primary outcome was specified as the percentage of responders at week 52 according to a novel composite disease activity measure (SRI). This outcome was statistically in favour of belimumab (10mg/kg vs. placebo) in both trials. For both whole and Target populations, results from BLISS-52 were more favourable for belimumab than results from BLISS-52.

For all secondary outcomes in BLISS-76 effect sizes were insufficient to be confident that effects could not be accounted for by chance. For several outcomes, including percentage responders and time to first flare in BLISS-76, although formal statistical tests were not performed, it appeared that the 1mg/kg dose regimen was as effective, or more effective, than the 10mg/kg dosage.

Geographical and racial differences between BLISS trials indicate that efficacy results from BLISS-76, rather than from BLISS-52 or pooled BLISS populations, are more generalisable to the UK.

On most safety outcomes placebo and belimumab performed equally. There were more deaths under belimumab than placebo; on a "per patient year of exposure" basis the rate for belimumab was about double that for placebo although this finding could have occurred by chance. Causes of death were various and most were those associated with the condition of SLE. There was insufficient evidence to determine if there was any association between belimumab and mortality.

5 ECONOMIC EVALUATION

5.1 Introduction

Including a one page summary of structure, assumptions and sources, with signposting to Tables.

Patient population and subgroups under consideration

The submission outlines that there are three groups under consideration:

- The patient population as observed from pooling the All BLISS patient data;
- The anticipated license patient population of Anti-dsDNA+ve and low (C3 or C4);
- The Target population which restricts the patient population to the licensed patient population with an SS score at baseline of at least 10.

With the exception of Table 6.49 of MS, the analysis presented within the main body of the submission relates to the All BLISS patient population. Little detail is presented for the anticipated license population, though the base case results for this group are presented within Table 6.49 of the MS. The inputs and results for the Target population are presented in section 6.9 of the MS.

Given the anticipated license as stated within the submission, the ERG review of the economics does not focus upon the All BLISS inputs, though the base case results for this group are presented. The brief summary of the base case results for the anticipated license patient population is also presented. But unless otherwise stated the inputs to and results of the modelling within the ERG review of the economics relate to the Target population.

Implementation of the electronic model

The manufacturer model is embedded within Excel, but apart from some very basic pre-modelling data adjustment the Excel element of this is confined to being a database of input values and a store of the model results. The modelling is implemented using Visual Basic (VB) programming.

Prior to running the model the user is allowed to change various pre-specified settings within the model, such as the subgroup to be analysed. The model uploads the relevant set of input parameters into memory, calculates the model using the VB code and outputs the results to an Excel worksheet.

The VB programming is well organised and compact with no obviously superfluous code. But it is relatively complex and lengthy with little to no explanatory comment, running to 53 Pages when printed out in Arial 8pt. This has made it difficult for the ERG to confidently explore structural scenarios other than those pre-specified within the model within the STA time constraints.

Stopping rule and clinical effectiveness estimates

Note that the economics applies a stopping rule within the belimumab arm at week 24: those not experiencing a change in their SELENA-SLEDAI (SS) score of at least 4 by week 24 are assumed to stop belimumab treatment. Conceptually, this gives rise to two groups within each arm:

- Belimumab week 24 responders;
- Belimumab week 24 non-responders;
- SoC week 24 responders;
- SoC week 24 non-responders.

When reviewing the economics of the submission, it should be borne in mind that the actual experience of the belimumab week 24 non-responders at week 52 as reported within the trials is not used within the model for these patients.

5.2 Manufacturer's submission

5.2.1 Economic literature search

The searches undertaken by the manufacturer to identify cost-effectiveness evidence were conducted on 18th March 2011. Seven databases were searched (Medline (Pubmed), Medline In-Process (Pubmed), Embase, EconLit, CRD databases (HTA database, Database of Abstracts of Reviews of Effectiveness (DARE), NHS Economic Evaluation Database (EED). In addition, searches were conducted in Research Papers in Economics (RePEc), a clinical trials database (ClinicalTrials), the websites of the American College of Rheumatology, the USA FDA and EMA.

The search utilised terms to identify patient group (lupus) and the intervention (belimumab). Terms to identify comparators were not included. In the Pubmed search, the restriction to title and abstract in the belimumab section of the search strategy has resulted in 7 fewer hits compared to the same line in the clinical effectiveness search strategy. In line 1 of the Pubmed search, lupus would automatically have been mapped by Pubmed, resulting in the inclusion of the wrong MeSH heading Lupus Vulgaris. The MeSH heading Lupus

Erythematosus, Systemic was not included. However, because Pubmed also searched for lupus in all fields, papers with this MeSH heading should still have been picked up.

In the Embase search several lines include major mistakes resulting in it being unclear as to how the database would have performed the search. For example, line 1 starts with "exp AND" and includes two database index terms that do not exist: 'lupus'/exp, 'sle'/exp. Testing this line of the search by entering it exactly as reported results in 23221 (06/05/11), which is far fewer than the 58059 reported in the search strategy. Many of the lines in Embase that should have been searching the EMTREE headings were entered very differently and resulted in far fewer hits (e.g. the EMTREE heading in the filter at line 11: exp Economic Evaluation/ (which when tested brought back 166263 hits on 06/05/11) was entered as "exp AND economic AND ('evaluation'/exp OR evaluation)" and resulted in only 977 hits, which may have led to important studies being missed.

No date or language limits were applied. A search filter was applied to the Embase and Medline searches to limit them to a particular type of evidence (economic studies), which was not appropriate in Pubmed in light of the small number of studies retrieved before the filter was applied (41). The manufacturer states that the search filter used was the CRD sensitive economics filter. However, the version the manufacturer uses in Pubmed does not match that given in CRD's NHS Economics Evaluation Database Handbook 2007[16] and there are also some mistakes in the translation of some elements (e.g. the MeSH heading in the filter exp "Costs and Cost analysis"/ was entered as: costs AND "cost analysis". Fortunately, when this was translated by Pubmed the correct MeSH heading was searched, but several unusual combinations of free text terms were also searched (e.g. costs"[All Fields] AND "cost"[All Fields] AND "analysis"[All Fields]).

Overall, the Pubmed and Embase searches appear to be of a poor quality and may have resulted in important studies being missed (see Appendix 2 for further details).

5.2.2 Manufacturer's direct drug cost and administration

Belimumab dose and direct drug cost

To calculate the direct drug cost for belimumab the manufacturer has assumed whole vial use that minimises wastage. Belimumab is available in 400mg and 120mg vials, at unit costs of £381 and £114 respectively excluding the PAS, these yielding the same cost per mg. [redacted] Since 120mg is not a factor of 400mg, the

minimum cost dose can in theory use anything up to nine 120mg vials. Cost minimisation suggests that where appropriate multiple vials of 120mg should be used. For instance, six 120mg vials provides a total available dose of 720mg whereas one 400mg vial and three 120mg vials provide a total available dose of 760mg: a 72kg patient is most cheaply dosed with six 120mg vials while a 73kg patient is most cheaply dosed with one 400mg vial and three 120mg vials.

The manufacturer applies the above dosing calculations to the distribution of patient weights of the pooled trial data to arrive at a weighted average belimumab drug cost. These are differentiated by patient subgroups to yield average drug costs per administration of £671 for the belimumab arm pooled for All BLISS, £650 when this is restricted to the anti-dsDNA+ve, low (C3 or C4) group and £654 when this is restricted to the Target population. [redacted]

Belimumab administration cost

Belimumab administration is assumed to require 2 hours of dedicated nursing time which is costed using 2010 Patient Social Service Research Unit (PSSRU) rates for a senior hospital staff nurse at £126 per administration. There is no specific allowance for any consumables within this administration costing

5.2.3 Model Structure

The model is implemented as a patient level simulation due to the complexity of SLE and the large number of health states that this implies. This inevitably makes it and its electronic implementation relatively complex, but an outline of the model and the data sources is reasonably simple to present. Within this summary it is simplest to separate the model elements by the source of the data feeding into them:

1. Trial data:
    a. The baseline characteristics for each patient being simulated including SS score at baseline and whether there is involvement for each of the 12 SLICC organs modelled
    b. The likelihood of response at week 24 in the belimumab arm, defined as a change of at least 4 in the SS score from baseline
    c. The change in SS score between baseline and week 52 for belimumab week 24 responders

   d. The change in SS score between baseline and week 52 for SoC, this also being applied to belimumab week 24 non-responders
   e. The "natural" discontinuation rates for belimumab week 24 responders after week 24
   f. The direct effect of the SS score upon quality of life
   g. The direct effect of the SS score upon treatment cost
2. JHU cohort data:
   a. The evolution of the SS score in the SoC arm after week 52, with belimumab week 24 responders being assumed to retain the absolute advantage in SS score over the SoC arm while they remain on treatment
   b. Implicit in the above the Adjusted Mean SLEDAI (AMS): the adjusted mean SS score since baseline
   c. The relationship between the SS score and steroid use
   d. The main survival model
   e. The likelihood of developing involvement for each of the 12 SLICC organs modelled if the organ concerned is not involved at baseline
3. Other data drawn from the broader literature:
   a. The standardised mortality rate for a given SS score
   b. The quality of life impact of organ involvement
   c. The additional cost for each organ involvement

The baseline characteristics, likelihood of response, week 52 SS scores and discontinuation rates are differentiated by subgroup within the model (1.a. - 1.e.). All other relationships are not. A number of Tables from the manufacturer's submission are replicated within what follows for ease of reference.

Trial data element 1.a.Baseline patient characteristics

For each patient level simulation the patient characteristics are randomly sampled from the underlying distribution; e.g. for the Target population the likelihood of the patient being female is based upon the 94.2% of the pooled trial data and a drawing on a Bernoulli distribution (Tables 6.37, 6.38 and 6.39 of the MS for the Target population). The patient is then cloned within the model for running through the SoC arm of the model and the belimumab arm of the model.

Trial data element 1.b.Likelihood of response in the belimumab arm

This is differentiated by the patient baseline SS score and drawn on a Bernoulli distribution (Table 6.42 of the submission for the Target population).

Trial data element 1.c and 1.d.Change in SS score from baseline to week 52

Unlike the likelihood of response, this is not drawn from lookup Tables based upon SS score at baseline and treatment arm. The manufacturer pools the trial data and uses regression analysis to derive coefficients for the percentage reduction in a patient's baseline SS score dependent upon whether the patient was in the SoC arm, the belimumab arm and if in the belimumab arm whether they were a week 24 responder (See Table 21; Adapted from Table 6.41 of MS for the Target population).

Table 21: Linear regression of coefficients for $SS_{52}=(1+\beta)SS_0$: Target population

	β	s.e.	P value
SoC	-34.9%	2.2%	< 0.01
Belimumab	-34.3%	4.6%	< 0.01
Belimumab week 24 responders	-28.0%	5.2%	< 0.01

As the model assumes that week 24 non-responders within the belimumab arm cease treatment and experience the SoC SS scores at week 52 the central estimates of this are more transparently rearranged as shown in Table 22.

Table 22: Rearranged linear regression of coefficients for $SS_{52}=(1+\beta)SS_0$: Target population

	β	1+β
SoC	-34.9%	65.1%
Belimumab week 24 non-responders	-34.9%	65.1%
Belimumab week 24 responders	-62.3%	37.7%

For a belimumab week 24 responder, the absolute difference at week 52 in SS scores between the belimumab week 24 responder and her SoC arm clone is assumed to be maintained while she remains on belimumab treatment; i.e. 27.4% of her baseline SS score up to discontinuation.

Trial data element 1.e. natural discontinuation rates for belimumab week 24 responders

The written submission is not entirely explicit as to the modelled natural discontinuation rates but it appears that for belimumab week 24 responders a daily hazard of discontinuation is calculated based upon the overall rate of discontinuations between day 168 and day 532. A six monthly natural discontinuation rate is calculated for the first year; i.e. presumably

subsequent to the assessment of response at week 24, with an annual rate being calculated thereafter. The day 168 to day 532 proportion remaining on treatment among belimumab week 24 responders is given as 0.891 in the electronic model for the All BLISS data set, and as 0.920 for the Target population.

A key aspect of this data is that it must relate to discontinuations between day 168 and day 532 and not to discontinuations between baseline and day 532, as the latter would cause the model to overestimate discontinuation rates among belimumab week 24 responders (Table 6.41 of the MS for the Target population). There is some lack of detail within the submission around this variable: its source, the period it relates to, any pooling of data between BLISS-52 and BLISS-76 given their differing duration after day 164; any evidence of difference between BLISS-52 and BLISS-76 for this variable; and, the reasons for the observed discontinuations. The latter may be particularly important given that this variable is used to extrapolate over the time horizon of the model with the cost effectiveness estimate being quite sensitive to it.

Within the model the impact of these continuation/discontinuation rates are graphed below, as submitted by the manufacturer in response to ERG clarification question B24. The upper curve is the Target population. Note that Figure 23 shows the effect of both discontinuations and mortality on those remaining on belimumab treatment.

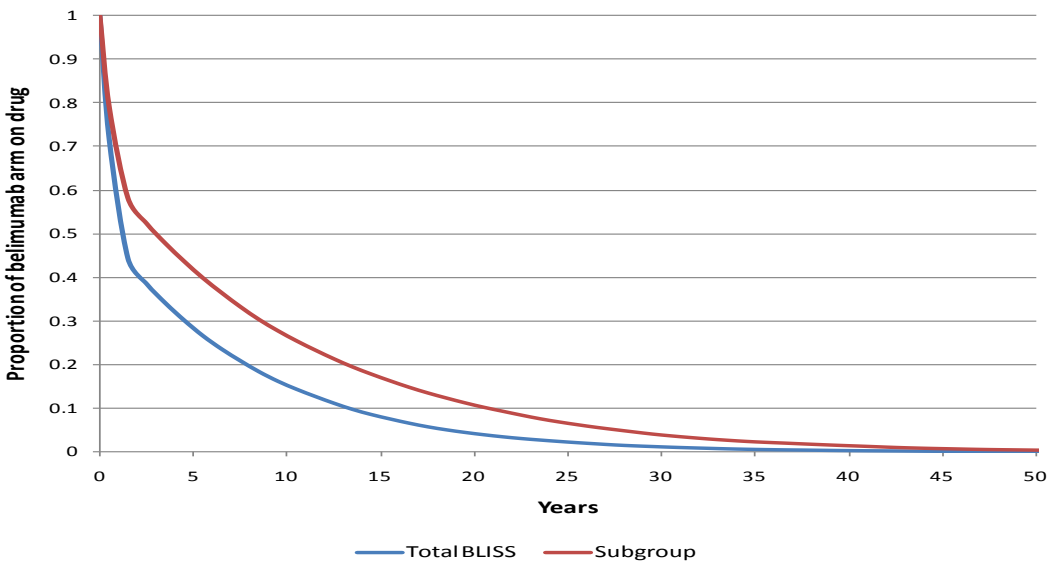


Figure 23: Continuation rates among belimumab week 24 responders

As explored later, a high discontinuation rate improves the estimated cost effectiveness of belimumab; i.e. it is more cost effective if belimumab has an initial effect and high response

rate but that patients experiencing a response rapidly discontinue belimumab treatment thereafter.

Trial data element 1.f. SS score direct impact upon QoL

Within the BLISS trials EQ-5D data was collected and transformed to HRQoL values using the Dolan algorithm[17]. Regression analysis then related these values to patients' SS scores, age, sex, ethnicity and organ damage. The statistically significant organ damage parameters were retained in order to control for their impact within the regression analysis, but organ damage was then set to be zero to generate the "clean" utility function for a patient with no organ damage:

"Clean" QoL = 1.297 – 0.145 * ln(Age) – 0.054 * ethnicity – 0.009 * SS score

where ethnicity is 1 if black and 0 if not. Note that the above corresponds with the clean utility function as in Table 6.20 of the submission and the electronic model, not with the function given in the text of the submission which appears to be incorrect.

Trial data element 1.g. SS score direct impact upon costs:

Limited detail is provided within the submission on the resource use questionnaire administered during the LBSL02 phase II trial. This was apparently a retrospective data collection administered at baseline, day 168 and day 365. 2006 PSSRU and NHS reference costs[18] were applied to this data, and the aggregate six monthly costs from baseline to day 168 and day 168 to day 365 were regressed on patients' SS score severity class during this period. The SS score severity class took a value of 0 to 3, this being determined by a patient's maximum observed SS score during the relevant 6 month period: 0 for a maximum SS score of 0; 1 for a maximum SS score of between 1 and 4; 2 for a maximum SS score of between 5 and 12; and, 3 for a maximum SS score of over 12. This regression analysis based on SS score severity class was then mapped back onto SS scores as outlined in Figure 6.16 of the submission, with the CPI being used to inflate the figures to 2010 prices and the six monthly costs being doubled to yield an annual cost. This yields the final direct cost function of Table 6.25 of the MS (Table 23). Within this, it should be noted that while both constant and derived coefficient were estimated as being significant, the regression had an R^2 of only 0.01.

Table 23: Manufacturer estimated SS direct annual cost function table

SS score	Cost	SS score	Cost	SS score	Cost
0	£1,152	5	£1,625	10	£1,931
1	£1,286	6	£1,681	11	£2,005

2	£1,419	7	£1,736	12	£2,079
3	£1,514	8	£1,792	13+	£2,153
4	£1,569	9	£1,857		

JHU cohort data elements 2.a.and 2.b. Evolution of SS and AMS scores after week 52

For the evolution of the SS score beyond week 52 the manufacturer originally estimated the regression equation:

But the manufacturer views this as providing a relatively poor fit to the data from the belimumab phase II trial, as shown in Figure 24. For the modelling the manufacturer adjusts the constant and applies the equation:

on the grounds of it better reflecting the phase II trial data.

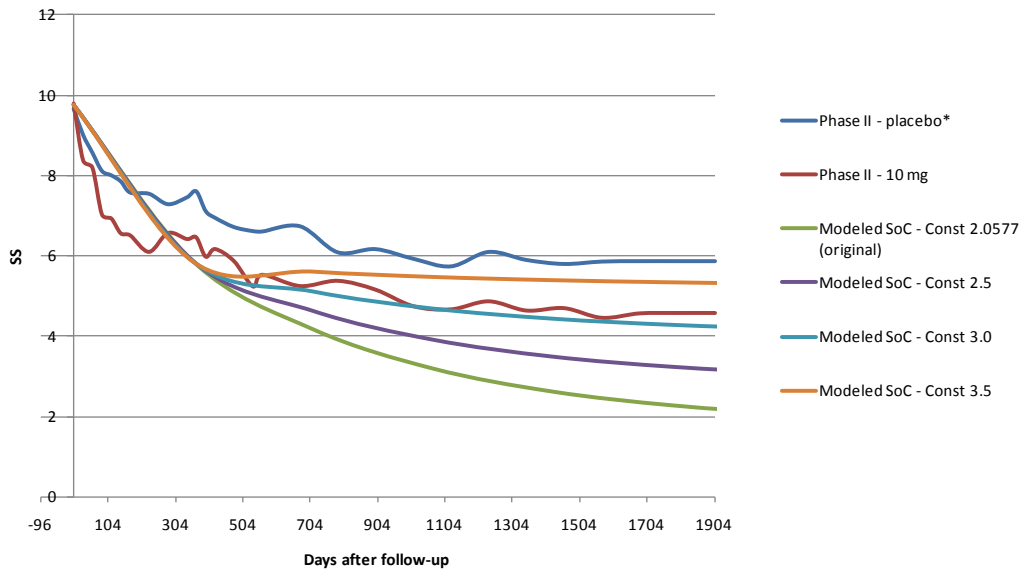


Figure 24: Medium term SS natural history model

The AMS score was developed to measure disease severity over time as opposed to the SS score which only reflects disease activity over the preceding 10 days. AMS is calculated as the area under the curve of disease activity measurements between two time-points. The area under the curve is then divided by time of follow-up to provide an average score over the period of interest. In Figure 24 above the ERG assume that the area under the curves shown can be used to represent AMS. However, the MS references to AMS score may refer to either the "AMS over lifetime" or the "average mean SLEDAI up to current time" which are presumably calculated in the same manner: the area under the SS score curve divided by time elapsed.

The AMS is calculated as the area under the SS score curve between two time points divided by the time of follow-up to provide an average score over the period of interest.

The AMS score refers to either the "AMS over lifetime" or the "average mean SLEDAI up to current time" which are presumably the same[a]: the area under the SS score curve divided by time elapsed.

The SS curve determines the AMS, with it being the AMS that is used within the survival model and the organ damage natural history model.

JHU cohort data element 2.c. Steroid use:

[a] Note that within the written there seems to be some occasional looseness of wording around SS and AMS, with there being some instances of the AMS referring to an annual AMS score; i.e. the average SS score over one year rather than from baseline.

The manufacturer argues that the steroid doses and changes to these as seen in the BLISS trials are not representative of the likely steroid dose reductions that would be possible with belimumab. Given this, the manufacturer fits a random effects model to the JHU data which estimates the steroid dose as a function of the average SS score within the year being simulated.

Table 24: Steroid use as a function of SS score Table 6.11

	Coef (95% CI)	P-value
Constant	3.410 (0.617-0.823)	< 0.001
SS score within year average	0.720 (3.073-3.747)	< 0.001

Note that despite the arguments around the representativeness of the JHU cohort for the All BLISS population and modelling as outlined in the previous section, no alternative forms for the steroid use equation were explored. In response to ERG clarification question C2 the manufacturer justifies this on the basis of there being little difference in the baseline steroid doses between the JHU cohort, 9.95mg/day, and the pooled All BLISS 10.78mg/day.

JHU cohort data element 2.d. Survival model

Initial univariate survival analysis within an exponential regression model framework found a range of variables within the JHU cohort data to be predictors of mortality, including age and duration of disease which were each statistically significant. Through a process of multivariate stepwise covariate selection a sensitivity analysis around the proper distribution (Exponential and Weibull) the range of variables included in the survival model were reduced. Within this process age was not included due to concerns around it having a high positive correlation with disease duration. Age at diagnosis and disease duration was chosen instead for this selection process. Within the multivariate stepwise covariate selection disease duration was further eliminated, leaving only age at diagnosis [Table 6.12 and Table 5 of Appendix 21[b]].

Note that both the AMS and the Cumulative Average Prednisone Dose (CAPD) up to current time are explanatory variables within the JHU cohort survival model. As outlined above, these are determined by a patient's SS curve.

[b] Table 5 of appendix 21 outlines that the model adopted, model 5, has the second highest AIC of the four models this is reported for.

Since age and disease duration have been eliminated from the JHU cohort survival model, the manufacturer reintroduces age to the calculation of mortality risk using an SLE Standard Mortality Ratio (SMR) differentiated by age group drawn from the Bernatsky 2006[19] reference coupled with general population mortality rates as outlined below.

JHU cohort data element 2.e.Risk of developing organ involvement

Through a similar analysis to the survival model, the manufacturer estimates individual risk equations for the development of individual SLICC item organ involvement [Table 6.14 and for more detail Appendix 21 of the MS]. These individual models are described within the model as "JHU – AMS forced in, involvement removed".

The AMS is an explanatory variable within the models of the risk of: CVD, gastrointestinal, musculoskeletal, neuropsychiatric, ocular, peripheral vascular, pulmonary, renal and skin involvement. But is not an explanatory variable within the models of the risk of: diabetes, malignancy or gonadal failure involvement.

The CAPD is an explanatory variable within the models of the risk of: CVD, diabetes, gastrointestinal, musculoskeletal, ocular, gonadal failure and skin involvement. But is not an explanatory variable within the models of the risk of: malignancy, neuropsychiatric, peripheral vascular, pulmonary or renal involvement.

Any new involvement of an organ is assumed to occur at the average SLICC score for that organ observed as observed across the JHU cohort. As the manufacturer notes, this will tend to overestimate the SLICC score for that organ when involvement occurs, but this bias is likely to wane as time and the model progresses. The net overall impact of the assumption of a constant SLICC score at the average of that observed in the JHU cohort for newly incident organ involvement is consequently ambiguous.

Literature element 3.a.SLE SMR by age group

Due to age not being within the JHU cohort derived survival model, the manufacturer uses a set of age dependent SMRs for SLE patients relative to the general population as derived from the Bernatsky 2006[19] reference: 19.2 age 16-24, 8.0 age 25-39, 3.7 age 40-59 and 1.4 age 60+.

To calculate the likelihood of a patient dying during a cycle the model first derives the probability of death for this patient from the JHU survival model. The probability of death from the JHU survival model for a patient at the average value of the covariates observed

within the JHU cohort is then calculated. Dividing the first by the second yields the patient's hazard of death compared to the "average" JHU cohort patient.

This hazard is then multiplied by the age dependent SLE SMR as drawn from the Bernatsky 2006[19] reference and the age dependent general population risk of mortality, with the derived mortality rate then being adjusted back to being a probability [see MS Table 6.13 and untitled Table immediately after for a worked example].

Literature element 3.b HRQoL impact of organ involvement

Utility values for each SLICC element were drawn where possible by the manufacturer from HTAs available on the NICE website.

Paralleling the assumption that the average SLICC score for new organ involvement would be the average observed across the JHU cohort, the weights attached to each SLICC element utility value are the proportion of those elements observed within the JHU cohort. The resulting weighted average is raised to the power of the average SLICC score for those with that organ involvement within the JHU cohort as given in Table 6.16 of the MS.

For instance, for the calculation of the pulmonary involvement HRQoL based upon the text of the submission is 0.70 in Table 16.19 of the MS (refer to Table 25).

Table 25: HRQoL calculation pulmonary involvement from Table 16.19

SLICC Element	HRQoL	JHU %	Weighted	JHU SLICC	Final
Pulmonary hypertension	0.61	33%	0.20		
Pulmonary fibrosis	0.73	42%	0.31		
Shrinking lung (Chest XRay)	1.00	2%	0.02		
Pleural fibrosis (Chest XRay)	1.00	20%	0.20		
Pulmonary infarction/resection	0.94	4%	0.04		
Average across pulmonary			0.77	1.31	0.70

These organ involvement HRQoL values are applied multiplicatively. For a patient having developed more than one SLICC organ involvement, only the lowest HRQoL multiplier is applied to the "clean" utility.

Literature element 3.c Cost impact of organ involvement

A similar approach is undertaken for the cost impacts of organ involvement as for the QoL impacts, only with the number of patients in the JHU cohort experiencing the individual elements among those having had an event within the organ class giving rise to the weight to apply. These weights can sum to more than one due to a patient being able to experience more than one event. As with the calculation of the quality of life impacts this will tend to overestimate costs in the incident year and early years after incidence.

As these cost elements are less well documented in the submission than the HRQoL elements the full set is outlined below, with more detail being available in Appendix 28 of the submission. There are some minor discrepancies between the figures in Table 26 and those given in Table 6.26 of the MS for reasons that are unclear, but these will not affect results.

Table 26: Average costs for organ involvement

	Unit Costs			Average total	
	Year 1	Year 2+	Weight	Year 1	Year 2+
Ocular				**£1,518**	**£17**
Cataract	£1,553		96%		
Retinal damage / optic atrophy	£103	£64	27%		
Neuropsychiatric				**£3,659**	**£1,131**
Cognitive impairment			24%		
OR major psychosis	£1,122	£1,122	8%		
Seizures requiring therapy for 6 months	£826		19%		
Cerebral vascular accident ever or resection excl mal.	£8,066	£2,266	38%		
Cerebral vascular accident ever or resection >1	£8,066	£2,266	2%		
Cranial or peripheral neuropathy			43%		
Transverse myelitis	£4,772	£2,386	4%		
Renal				**£1,746**	**£2453**
				To max	**£6479**
Pulmonary				**£9,678**	**£9,603**
Pulmonary hypertension	£22,488	£22,488	43%		
Pulmonary fibrosis			55%		
Shrinking lung (on chest radiograph)			3%		
Pleural fibrosis (on chest radiograph)			26%		
Pulmonary infarction or resection	£1,539		5%		
Cardiovascular				**£3,402**	**£500**
Angina or Coronary Artery Bypass Graft	£4,196	£368	31%		
Myocardial infarction	£4,322	£368	36%		
Cardiomyopathy	£724	£724	35%		
Valvular disease (dias/sys murmur)			25%		
Pericarditis x 6 mth or pericardiectomy	£2,079		14%		
Peripheral vascular				**£2,955**	**£591**
Significant tissue loss ever	£10,375	£368	21%		
Significant tissue loss > 1 site			0%		
Venous thrombosis with swelling	£1,501	£936	55%		
Gastrointestinal				**£2678**	**£0**
Infarction or resection of bowel	£2,848		93%		
Resection > 1 site	£2,848		1%		
Pancreatic insufficiency enzyme replacement			3%		
Musculoskeletal				**£5,372**	**£1,903**
				To min	**£1,514**
Muscle atrophy / weakness			11%		
Deforming or erosive arthritis	£3,112	£3,112	26%		
Osteoporosis with fracture or vert. collapse	£8,118	£1,148	49%		
Avascular necrosis	£1,359		37%		
Avascular necrosis 2	£1,359		3%		
Osteomyelitis			2%		
Ruptured tendon			12%		
Diabetes				**£2,313**	**£2,313**
Diabetes mellitus sufficient for some intervention	£2,313	£2,313	100%		
Malignancy				**£6,056**	**£0**
Malignant tumours	£6,056		100%		

5.2.4 Base case deterministic results

The base case deterministic results are presented below in Table 27, Table 28, Table 29, Table 30 and Table 31.

Both Table 5 and 6 of the MS and also the default belimumab costs in the electronic model, suggest that the [redacted]

Table 27: Base case deterministic results: All BLISS

		Without PAS		With PAS	
	SoC	Belimumab	Net	Belimumab	Net
Undiscounted survival Life Years	30.47	31.97	1.50	31.97	1.50
Discounted quantities					
Belimumab direct drug cost	..	£31,687	£31,687	[redacted]	[redacted]
Total cost	£97,583	£133,167	£35,584	[redacted]	[redacted]
QALYs	9.55	9.98	0.43	9.98	0.43
Base Case ICER			£82,909		[redacted]

Table 28: Base case deterministic results: Anticipated license population

		Without PAS		With PAS	
	SoC	Belimumab	Net	Belimumab	Net
Undiscounted survival Life Years	32.82	34.96	2.13	34.96	2.13
Discounted quantities					
Belimumab direct drug cost		£36,796	£36,796	[redacted]	[redacted]
Total cost	£103,591	£143,895	£40,303	[redacted]	[redacted]
QALYs	10.11	10.72	0.61	10.72	0.61
Base Case ICER			£66,170		[redacted]

Table 29: Base case deterministic results: Target population

		Without PAS		With PAS	
	SoC	Belimumab	Net	Belimumab	Net
Undiscounted survival Life Years	31.93	34.87	2.93	34.87	2.93
Discounted quantities					
Belimumab direct drug cost	0	£47,008	£47,008	[redacted]	[redacted]
Total cost	£105,366	£157,291	£51,925	[redacted]	[redacted]
QALYs	9.81	10.61	0.81	10.61	0.81
Base Case ICER			£64,410		[redacted]

From the above the difference between the estimates of cost effectiveness for the anticipated license population and the Target population are relatively minor. Given this, from an economic point of view it is unclear why the manufacturer niches belimumab to only those with an SS score of at least 10 at baseline: around 67% (n=396) of the anticipated license population (n=592) within the trials.

Table 30: Base case organ involvement to death MS Table 6.43: Target population

	SoC	Belimumab	Net
Cardiovascular	23.9%	21.3%	-2.6%
Diabetes	17.9%	19.2%	1.3%
Gastrointestinal	22.1%	25.0%	3.0%
Malignancy	32.0%	34.1%	2.2%
Musculoskeletal	48.5%	48.9%	0.4%
Neuropsychiatric	44.7%	45.8%	1.1%
Ocular	35.1%	36.0%	0.8%
Peripheral vascular	21.5%	20.8%	-0.7%
Premature gonadal failure	7.2%	7.2%	0.0%
Pulmonary	39.9%	36.8%	-3.1%
Renal	24.3%	19.2%	-5.1%
Skin	7.9%	7.9%	0.0%

Table 31: Base case discounted costs: Target population

	SoC	Belimumab	Net
Belimumab therapy	..	£47,008	£47,008
Belimumab administration	..	£9,059	£9,059
Belimumab total costs	..	£56,067	£56,067
Other Costs			
SS score related costs	£27,882	£28,130	£248
Organ damage costs			
Cardiovascular	£1,838	£1,633	-£205
Diabetes	£2,493	£2,731	£238
Gastrointestinal	£359	£399	£40
Malignancy	£998	£1,031	£33
Musculoskeletal	£9,758	£10,114	£356
Neuropsychiatric	£6,434	£6,719	£286
Ocular	£392	£391	-£1
Peripheral vascular	£1,380	£1,339	-£41
Premature gonadal failure	£0	£0	£0
Pulmonary	£42,692	£39,652	-£3,040
Renal	£11,139	£9,083	-£2,056
Skin	£0	£0	£0
Total other costs	£105,366	£101,224	-£4,142
Total costs	£105,366	£157,291	£51,925

As outlined in Table 31 the direct drug costs of belimumab account for around 90% of the total net costs with the administration costs of belimumab accounting for another 17% of the total net costs: taken together roughly 108% of the total net cost. Administration costs are a significant proportion of the total direct cost of belimumab: 16% without the PAS and approximately [redacted] with the PAS.

The main anticipated cost savings arise from reduced rates of pulmonary disease and renal involvement, which generate cost offsets of around -6% and -4% of the total net costs respectively. Some additional costs are associated with belimumab due to the anticipated undiscounted survival gain of 2.93 life years causing some complications to occur in a higher proportion of patients and to be experienced for longer.

5.2.5 Base case probabilistic results

The base case probabilistic results do not appear to be presented within the written submission, with only the CEACs being presented. Refer to Figure 25, Figure 26 and Figure 27.

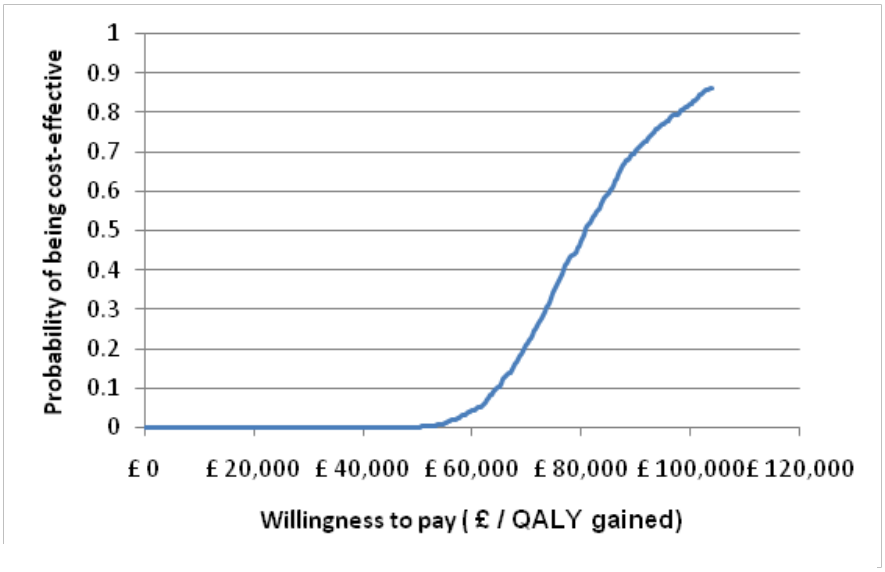


Figure 25: CEAC without PAS (MS Fig 6.27) All BLISS

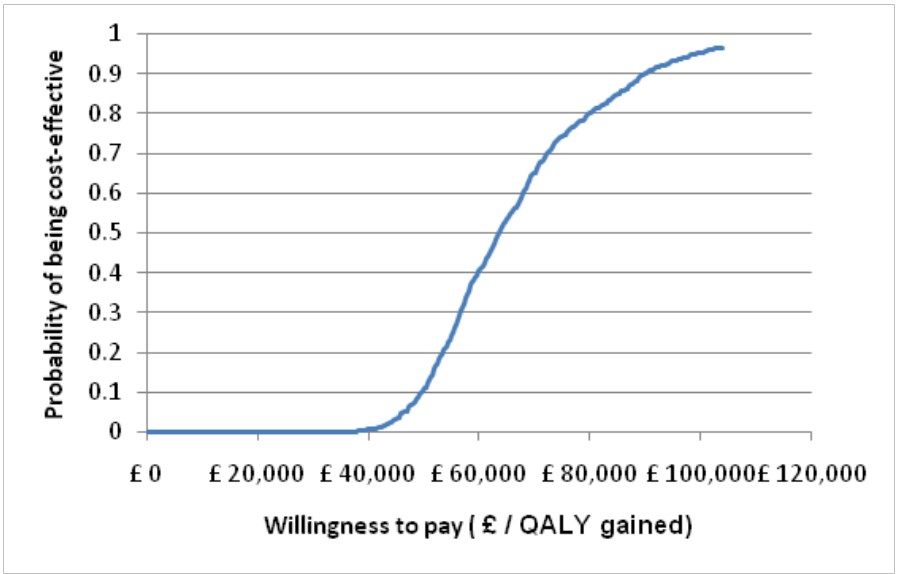


Figure 26: CEAC without PAS (MS Fig 6.41) Target population

[redacted]

Figure 27: CEAC with PAS, corrected in response to clarification question: Target population

[redacted]

Manufacturer's sensitivity and scenario analysis

The presentation of the sensitivity analyses within the submission and the responses to ERG clarification questions are mainly limited to those for the Target population. The results of sensitivity analyses presented by the manufacturer for other groups are only presented here when that for the Target population is not supplied by the manufacturer.

As outlined in the summary of the model structure for the evolution of SS scores the manufacturer estimated a regression from the JHU cohort that related the change in the SS score to the average SS score in the previous period, and to gender, ethnicity and age. The constant estimated from this regression was 2.0577. The modelling applied a value of 3.0 on the basis of 2.0577 providing a poor fit to the evolution of SS scores within the All BLISS data. The effect of this upon the cost effectiveness for the All BLISS group is as in Table 32.

Table 32: Varying the constant in the SS change regression: All BLISS

Regression constant	ICER
2.0577	£93,654
2.5	£85,394
3.0	£82,909
3.5	£80,988

The univariate sensitivity analyses undertaken by the manufacturer are presented graphically as a tornado diagram in Figure 6.37 of the MS, with the values underlying this shown in Table 33. These values relate to the without PAS scenario for which the base case Incremental Cost-Effectiveness Ratio (ICER) is estimated as £64,410 per QALY.

Table 33: Manufacturer univariate sensitivity analysis – Target population

	Base	Low		High	
Variable	**Value**	**Value**	**ICER**	**Value**	**ICER**
Belimumab wk 24 responders % SS change baseline to 52 weeks	-0.28	-0.383	£49,393	-0.173	£103,840
All belimumab patients % SS change baseline to 52 weeks	-0.343	-0.437	£50,335	-0.251	£96,031
AMS on mortality	0.213	0.085	£85,677	0.333	£50,962
Annual % belimumab week 24 responders on treatment	0.92	0.863	£54,518	0.981	£85,893
Ln(age) of the "clean utility" regression	-0.145	-0.18	£78,448	-0.103	£53,263
Constant coefficient in "clean utility" regression	1.297	1.146	£79,243	1.426	£55,493
SoC patients % SS change baseline to 52 weeks	-0.349	-0.394	£77,351	-0.307	£55,581
AMS of the natural history Pulmonary model	0.139	0.06	£73,044	0.216	£55,216
Constant of the natural history Neuropsychiatric model	-7.396	-9.934	£61,333	-5.117	£76,231
Ln(age) of natural history Neuropsychiatric model	0.607	0.026	£61,514	1.226	£76,261
AMS coefficient of the natural history Renal model	0.323	0.228	£69,696	0.412	£56,744
Constant of the adjusted natural history SS model	3	2.202	£73,226	3.934	£61,871
Constant coefficient from the natural history PV model	-11.69	-16.47	£65,935	-6.81	£55,396
Ln(age) of natural history Pulmonary model	1.232	0.593	£70,841	1.916	£79,571
Constant of the natural history Renal model	-8.29	-9.01	£67,867	-7.56	£60,057

Many of the key variables within the sensitivity analyses tornado diagram are as would be expected: including the changes in SS scores, and the impact of SS scores through the AMS upon mortality,

As for the All BLISS sensitivity analysis the adjustment to the constant for the adjusted natural history model of the evolution of the SS score, demonstrates a similar impact as in the model for the Target population. The value of 2.202 (close to the original estimate of 2.0577) gives a cost effectiveness estimate of £73,226 per QALY. The alternative estimate of 3.9 results in a cost effectiveness estimate of £61,871 per QALY. One point to note is that this is

non-linear: an increase of 0.934 improves the cost effectiveness estimate by £2,539 per QALY while a smaller reduction of 0.798 worsens the cost effectiveness estimate by £8,816 per QALY.

Given the cost outputs of the model as previously summarised, the specifications of the natural history models for pulmonary and renal disease also have an impact. Perhaps more surprising is the influence of the neuropsychiatric natural history model.

Note that the manufacturer has presented additional information in Tables B8.2 and B8.3 in the response to ERG clarification questions. These outline the sensitivity of net QALYs and net costs to the univariate sensitivity analyses. The neuropsychiatric natural history model mainly impacts upon QALYs. Within net costs, the renal natural history model also has an impact.

A key variable that has not been particularly explored or explained within the submission is the assumed rate of continuation and discontinuations among belimumab week 24 responders. The ICERs reported in MS provide insufficient detail. Table 34 shows that both the net QALYs and the net cost are increasing in the annual continuation rate for belimumab week 24 responders, as would be anticipated. Note that the longer belimumab week 24 responders are estimated to remain on belimumab treatment the worse the estimated cost effectiveness is for belimumab.

Table 34: Sensitivity to continuation rate for belimumab week 24 responders: Target population

	Low value	Central value	High value
Annual continuation rate	0.863	0.920	0.981
Net QALY	0.649	0.806	1.165
Net Cost	£35,386	£51,925	£100,094
ICER	£54,518	£64,410	£85,893

The manufacturer also presented a range of scenario analyses for the Target population for the without PAS scenario. Compared to the base case estimate of £64,410 per QALY:

- Excluding the continuation rule at week 24 worsens the ICER to £72,207 per QALY
- Tightening the continuation rule to SS change≥ 6 improves the ICER to £50,114 per QALY
- A 12% higher belimumab price worsens the cost effectiveness to £71,297 per QALY
- Vial sharing improves the cost effectiveness to £61,671 per QALY

- Using the original natural history model worsens the cost effectiveness to £77,707 per QALY
- An administration cost of £159 worsens the cost effectiveness to £67,353 per QALY

5.2.6 Base case deterministic results

Base case results

Simulating 50,000 patients within the deterministic model results in mean estimates that cross check with those presented within the submission.

5.2.7 Base case probabilistic results

Due to the patient level simulations for the base case deterministic results being run across 50,000 simulations, running the model for a sufficiently large number of iterations for the probabilistic modelling to reliably generate a central estimate and associated CEAC takes around one week. Time constraints have precluded the ERG from running the model probabilistically for all the subgroups and for any of the sensitivity analyses.

The results of an ERG probabilistic run of the model for the Target patient population, retaining 50,000 patients over 1,000 iterations, cross check with that of the manufacturer, the likelihood of belimumab being cost effective for a given willingness to pay (WTP):

- 0% at £30,000 per QALY
- 1% at £40,000 per QALY
- 11% at £50,000 per QALY
- 40% at £60,000 per QALY

In addition, the central estimates from the ERG probabilistic run of the model are an additional 0.84 QALYs at an additional cost of £55,166 to yield a central estimate of cost effectiveness of £65,530 per QALY. Both net QALYs and net costs are slightly higher than the deterministic run of the model, 0.81 and £51,925 respectively, but both rise in roughly equal proportion and the central estimate from the probabilistic modelling of £65,530 per QALY is little different from the deterministic estimate of £64,410 per QALY.

5.2.8 Data inputs

For correspondence between written submission and electronic model related to the Target population please refer to Table 35.

Table 35: Correspondence between MS and electronic model: Target population

Written submission		Electronic model		Correspondence
Model parameters specific to Target population				
Data	Table	Worksheet	Cells	Correspondence
Baseline demographics	6.37	Baseline Patient Characteristics Subgroup BLISS data	D9 Q7:Q62, P64, P78, P69, P240	Yes
Baseline disease activity	6.38	Baseline Patient Characteristics Subgroup BLISS data	D21:D32 Q185:Q215 P258:Q258 P260:Q260 P262:Q262 P265:Q265 P267:Q267 P269:Q269 P271:Q271 P273:Q273	Yes
Baseline SDI items	6.39	Subgroup BLISS data	P245:T256	Yes
Discontinuation rates	6.40	Treatment Effect Baseline Patient Characteristics Subgroup BLISS data	L33:L34 AD38 Q185:Q215 AR9:AR39 P276:Q276	
Week 52 SS regression	6.41	Treatment Effect	D12:E14 P221:Q223	Yes
Week 24 response by SS	6.42	Baseline Patient Characteristics Subgroup BLISS data	AD7:AD38AR9:AR39	Yes
Model parameters common to all subgroups				
Data	Table	Worksheet	Cells	Correspondence
Long term SS regression	6.9	Natural History Short	D11:D16	As per the text of the submission, the constant within Table 6.9 is increased from 2.0577 to 3.0000. The electronic model corresponds with the constant being 3.0000
Steroid use	6.11	PSA Inputs	HY7:HZ7	Yes
Mortality weibull	6.12	Natural History Model data	AG8:AG52	Yes
SLE SMR by age	6.13	PSA Inputs	IB7:IE7	Yes
Organ damage tte	6.14	Natural History Model data	AI8:AT52	Yes
JHU cohort characteristics	6.15	Hopkins Patient Characteristics	D8:D38	Possibly not. The electronic model highlights a larger range of cells in white than the five rows within Table 6.15: 19 cells in total. These cells are described within the electronic model as "*Average characteristics imputed to simulate non-trial*

Written submission		Electronic model		Correspondence
				characteristics that are used to determine long term outcome risks". Some of these elements of the electronic model may relate to those given in Table 6.16 of the submission Baseline hypertension within the electronic model is given as 53.1% compared to the 15.8% annual risk within Table 6.15. Note that the electronic model given the annual infection probability as 15.8%
JHU SLICC scores	6.16	Hopkins Characteristics	D8:D38?	Possibly not. There is no ready read across between D8:D38 of Hopkins Characteristics and Table 6.16. For instance, the electronic model give a value of 9.7% renal damage (mean) while Table 6.16 gives a renal score of 1.83. Even if the renal damage among the 9.7% had been at the maximum SLICC damage level of 4 it is difficult to see how this can result in a renal score of 1.83
Organ HRQoL impact	6.19	QoL Inputs	X9:AI10	Yes
"Clean" utility equation	6.19	QoL Inputs	D10:D13	No. The electronic model applies the values given in Table 6.20 and not those given in Table 6.19 and in the body of the text of the submission
"Clean" utility equation	6.20	QoL Inputs	D10:D13	Yes
Cost for a given SS score	6.25	Other cost inputs	D9:D29	Yes
Organ damage cost	6.26	Other cost inputs	W8:AH9	Partial. The year 1 and year 2 costs in the electronic model are as per Table 6.26 But note that the musculoskeletal annual cost declines between year 2 and year 17 from £1,903 to £1514 Also note that the renal annual cost increases

Written submission		Electronic model		Correspondence
				indefinitely from year 2 onwards from £2,453 to £6,749 by year 50
Belimumab annual costs	6.27	Belimumab cost	I23:I38	Yes

5.2.8.1 Model structure

Belimumab dose and direct drug cost

There appear to be minor errors in the calculation of the average belimumab drug cost per administration. For instance, for a patient weight of 50kg the manufacturer calculates that this is most cheaply administered using five 120mg vials at a cost of £571 per administration. This results in a total available dose of 600mg and wastage of 100mg, when a combination of one 400mg and one 120mg dose results in a total available dose of 520mg and wastage of only 20mg.

Note that if the simpler approach of using 400mg vials being used for the dose up to a multiple of 400mg with anything in addition to this being topped up through use of 120mg vials, wastage and the average drug cost would increase.

The drug cost for a patient of the mean patient weight can be calculated on the same basis as the manufacturer uses for individual patient drug cost calculations. Within an individual patient simulation model the approach of the manufacturer is correct. But much of the modelling submitted to NICE employs Markov modelling of a cohort of representative patients. This may use a weighted average drug cost, but it is also not unknown for the drug cost for the representative patient to be calculated. See Table 36.

Table 36: Belimumab average direct cost per administration

	Manufacturer	ERG	400mg	Mean		Median	
All BLISS	£671	£664	£694	67kg	£686	63kg	£610
Anti-dsDNA+ve, low (C3/C4)	£650	£642	£671	65kg	£686	62kg	£610
Target population	£654	£646	£674	65kg	£686	63kg	£610

Any errors within the manufacturer calculations are slight and will not materially affected the estimates of cost effectiveness. For the most severe subgroup of patients adopting the approach of using 400mg and topping up with 120mg rather than minimising waste has a similar effect to costing at the mean patient weight, with both increasing the average drug cost by a little over 3%.

As the belimumab drug costs account for around 90% of the estimated total net cost for both the All BLISS population and the most severe subgroup with an SS ≥ 10 at baseline, any change in the belimumab drug costs will lead to a roughly proportionate change in the cost effectiveness estimate.

Modelling the evolution of SS scores during the 1st year

The likelihood of the patient dying or developing organ involvement is not directly determined by the SS score, but this flow through to the AMS score which does determine mortality and organ involvement. The SS score has direct impacts upon patient utilities and the costs of treatment. The SS score also determines the patient steroid use, which in turn further affects the likelihood of mortality and organ involvement. The SS score is the key variable within the modelling.

The manufacturer response to ERG clarification question B3 outlines that the change in the SS score from baseline to week 52, SS_{52}-SS_0, for belimumab week 24 non-responders is calculated using the week 52 linear regression for the SoC arm. The reason for this is that belimumab week 24 non-responders cease belimumab treatment at week 24 and as a consequence at week 52 are receiving only SoC.

But calculating SS_{52} for belimumab week 24 non-responders using the week 52 linear regression for the SoC arm is likely to be incorrect. Both the SoC arm and the belimumab arm of the trials had week 24 responders and week 24 non-responders. For the Target population these are shown in Table 37 below.[c]

Table 37: Week 24 response rates – Target population

	BLISS-52		BLISS-76		Pooled	
	Resp	NResp	Resp	NResp	Resp	NResp
SoC	59%	41%	44%	56%	52%	48%
Belimumab	71%	29%	62%	38%	67%	33%
OR	1.75		2.07		1.93	

By definition the 52% of patients within the pooled SoC arm with response at week 24 had a change of at least 4 in their SS score at week 24, while the 48% without response at week 24 had a change of less than 4 in their SS score at week 24. Similarly by definition, the 33% within the pooled belimumab arm without response at week 24 had a change of less than 4 in their SS score at week 24.

[c] Based upon the patient numbers reported in table A7.1 of the manufacturer response to ERG clarification question A7.

It seems likely to be more appropriate to model SS_{52} for the belimumab week 24 non-responders based upon the changes in SS scores among the SoC week 24 non-responders than upon changes in SS scores across all SoC patients, the latter being an average across week 24 responders and week 24 non-responders. This is underlined by the response of the manufacturer to ERG clarification question A7 and associated Table A7.1 outlining the mean changes in SS scores at week 24 and week 52 by arm and week 24 responder status. See Table 38.

Table 38: SS changes at week 24 and by week 52 by week 24 status – Target population

	BLISS-52			BLISS-76			Pooled		
Week 24 status	Resp	NResp	All	Resp	NResp	All	Resp	NResp	All
Mean change from baseline at week 24: SS_{24}-SS_0									
SoC	-6.7	-0.7	-4.2	-7.1	-1.3	-3.9	-6.9	-1.1	-4.1
Belimumab	-7.4	-0.8	-5.5	-7.0	-1.0	-4.7	-7.3	-0.9	-5.2
Mean change from baseline at week 52: SS_{52}-SS_0									
SoC	-5.4	-2.3	-4.1	-6.1	-2.3	-4.0	-5.7	-2.3	-4.1
Belimumab	-7.5	-3.3	-6.3	-6.8	-2.5	-5.2	-7.2	-2.9	-5.8

Among week 24 non-responders the average changes in SS score at week 24 were very similar between the SoC arms and the belimumab arms: for instance these were -1.1 and -0.9 respectively within the All BLISS data.

It could be argued that among week 24 non-responders the average changes in SS score have started to slightly divergence between the SoC arms and the belimumab arms by week 52: -2.3 and -2.9 respectively within the All BLISS data. But for modelling purposes these figures are not particularly relevant. The stopping rule of the modelling is applied at 24 weeks. As a consequence, the evolution of SS scores among week 24 non-responders in the belimumab arm between week 24 and week 52 within the trials is of less interest. It reflects continued use of belimumab between week 24 and week 52 during the trials, when the modelling assumes that these patients will no longer receive belimumab between week 24 and week 52. It is presumably for this reason that SS_{52} has to be modelled for belimumab week 24 non-responders rather than drawn directly from trial data.

In the light of the above (see Table 38), for the belimumab week 24 non-responders it would seem to have been more appropriate to base SS_{52}-SS_0 upon the parallel change for SoC week 24 non-responders than that for the SoC arm as a whole: -2.3 rather than -4.1 within the All BLISS data. This is quite a large difference of -1.8 for belimumab week 24 non-responders given the overall changes in average SS scores within the trials. Belimumab week 24 non-responders make up 33% of the belimumab Target population within the trials.

The trial data reported in Table 39 is illustrative of the assumption underlying the manufacturer model, with the figures relating to trial data. The model makes a parallel assumption, but estimating SS_{52}-SS_0 uses the linear regressions as outlined in Table 6.41 of the MS and the manufacturer response to question C6 of the ERG clarification questions (see Table 39).

Table 39: Linear regression of SS_{52}-SS_0 central parameter estimates – Target population

	BLISS-52	BLISS-76	Pooled
SS_{52}-SS_0 SoC	-0.3629	-0.3341	-0.3493
SS_{52}-SS_0 belimumab	-0.3746	-0.3153	-0.3435
SS_{52}-SS_0 belimumab week 24 responders	-0.2626	-0.2827	-0.2800

Where SS_{52}-SS_0 SoC is the average SS change within the SoC arm, SS_{52}-SS_0 belimumab is the average SS change within the belimumab arm and SS_{52}-SS_0 belimumab week 24 is the additional average SS change within the belimumab arm among those showing a response at week 24. Note that these are regression coefficients and multiplicative: e.g. from the All BLISS data a patient within the SoC arm with, for example, SS_0 = 10.00 has a central estimate of SS_{52}-SS_0 = -0.349*10 = -3.49 hence SS_{52} = 6.51.

Note that the data in Table 39 will include any additional treatment effect from belimumab between week 24 and week 52 among belimumab week 24 non-responders. To the extent that this effect exists it will tend to lead to an overestimate of the effectiveness of for SS_{52}-SS_0 belimumab within the regression model, but this will tend to net out through a reduction in the estimate of SS_{52}-SS_0 belimumab week 24 responders. Within the modelling it is only really the sum of these two coefficients that it applied.

As outlined within the manufacturer response to the ERG clarification question B3, SoC patients are assumed to have the SoC coefficient applied, belimumab week 24 non-responders are assumed to have the SoC coefficient applied and belimumab week 24 responders have the sum of the two belimumab coefficients applied. Given this the model application of the results of the regression can be more transparently presented as below. These are then applied to the trial mean of $SS_0 \approx 12.7$ among the Target population to result in a modelled estimate for the trial mean which can then be compared with the actual trial mean for some simple triangulation and model validation.

Table 40: SS_{52}-SS_0 model versus trial – Target population

All BLISS	Coeff.	Modelled SS_0=12.7	Trial mean
SS_{52}-SS_0 SoC week 24 non-responders	-0.349	-4.4	-2.3
SS_{52}-SS_0 SoC week 24 responders	-0.349	-4.4	-5.7
SS_{52}-SS_0 SoC All (weighted average)		-4.4	-4.1
SS_{52}-SS_0 belimumab week 24 non-responders	-0.349	-4.4	-2.9
SS_{52}-SS_0 belimumab week 24 responders	-0.623	-7.9	-7.2
SS_{52}-SS_0 belimumab All (weighted average)		-6.8	-5.8

As anticipated, data shown in Table 40 suggests that assuming that belimumab week 24 non-responders have the same change in SS scores as the average for the SoC arm systematically overestimates the average change in the SS score within belimumab arm. Note that this also does not correct for any impact of belimumab week 24 non-responders being assumed to cease treatment at week 24, which may suggest that the -2.9 trial mean will be an overestimate of the likely effect in clinical practise and the value that should be applied within the modelling.

Note also that the electronic copy of the model includes the coefficients for the parallel regression of the change in SS scores at week 24: SS_{24}-SS_0

Table 41 mirrors the week 24 trial data, but appears unduly pessimistic for the belimumab week 24 non-responders when coupled with the average baseline of $SS_0 \approx 12.7$. This regression is not used within the current model implementation and is not presented within the written submission.

Table 41: Linear regression of SS_{24}-SS_0 central parameter estimates – Target population

	BLISS-52	BLISS-76	Pooled
SS_{24}-SS_0 SoC	n.a.	n.a.	-0.3525
SS_{24}-SS_0 belimumab	n.a.	n.a.	-0.0003
SS_{24}-SS_0 belimumab week 24 responders	n.a.	n.a.	0.5755

The manufacturer in response to ERG clarification question C8 confirms that the central estimates of the modelled evolution of SS scores for SoC, belimumab week 24 non-

responders and belimumab week 24 responders for the first 10 years[d] of the modelling are shown in Table 42.

Table 42: Modeled evolution of SS scores – Target population

Year	SoC	Bel. week 24 non-responders	Bel. week 24 responders
0	12.7	12.7	12.7
1	8.2	8.1	4.9
2	6.5	6.4	3.4
3	5.5	5.5	2.6
4	4.9	4.9	2.3
5	4.6	4.6	2.2
6	4.4	4.4	2.2
7	4.3	4.3	2.2
8	4.2	4.2	2.3
9	4.1	4.1	2.4
10	4.1	4.1	2.5

On the basis of a 67:33 split between belimumab 24 week responders and non-responders, the weighted average SS score within the belimumab arm is modelled as being 5.95 at the end of year 1: a net gain over SoC of around 2.2. This compares with a net observed gain from belimumab over SoC in the trials of 1.7: the modelled net improvement is around 30% greater than that observed in the trials. If the SoC week 24 non-responder change of -2.3 at week 52 is assumed to apply to the belimumab week 24 non-responder at week 52 on the grounds that these patients are assumed to cease belimumab treatment from week 24 the net "observed" gain from belimumab over SoC in the trials falls to 1.5: the modelled net improvement consequently increases to around 45% over that "observed" in the trials.

In short, given the trial data for the average changes in SS scores for SoC and belimumab week 24 responders and non-responders, it seems unwarranted to assume that belimumab week 24 non-responders will have the same change in SS score at week 52 as the average observed across the SoC arm. This overestimates the impact of belimumab upon SS scores.

Modelling of AMS scores

[d] For all the SS and AMS elements only the first 10 years modelling is presented for illustrative purposes, while the modelling extends to the patient lifetime.

It is not the SS score but the AMS score that directly contributes to the likelihood of a patient dying and a patient developing cardiovascular, gastrointestinal, musculoskeletal, neuropsychiatric, ocular, peripheral vascular, pulmonary, renal and/or skin involvement. Renal involvement further determines the likelihood of cardiovascular involvement.

As explained previously the AMS is calculated as the area under the SS score curve between two time points divided by the time of follow-up to provide an average score over the period of interest. Within the response to ERG clarification question C8 the manufacturer clarified the central estimates of the SS scores and AMS as below in Table 43.

Table 43: Modeled evolution of AMS scores: manufacturer clarification – Target population

Year	SoC		Bel. week 24 non-responders		Bel. week 24 responders	
	SS	AMS	SS	AMS	SS	AMS
0	12.7	12.7	12.7	12.7	12.7	12.7
1	8.2	10.4	8.1	10.4	4.9	9.9
2	6.5	9.1	6.4	9.1	3.4	8.4
3	5.5	8.2	5.5	8.2	2.6	7.4
4	4.9	7.6	4.9	7.5	2.3	6.8
5	4.6	7.1	4.6	7.0	2.2	6.3
6	4.4	6.7	4.4	6.7	2.2	5.9
7	4.3	6.4	4.3	6.4	2.2	5.6
8	4.2	6.1	4.2	6.1	2.3	5.4
9	4.1	5.9	4.1	5.9	2.4	5.2
10	4.1	5.8	4.1	5.8	2.5	5.0

Running the model with a 100% probability of response for belimumab and a 0% probability of response[e] for belimumab results in the same SS scores as above for SoC, Belimumab week 24 non-responders and belimumab week 24 responders.

For both the SoC arm and the belimumab week 24 non-responders, the AMS at time T as reported in the response to clarification question C8 cross checks with being the average of

[e] Implemented by setting the *Baseline Patient Chars* worksheet AD7:AD38 = 1 for 100% response and = 0 for 0% response and taking the values from the *Results* worksheet columns BR:BS.

the SS values t = 0…..T. But for the belimumab week 24 responders the AMS at time T as reported in the response to clarification question C8 bears no relation to the ERG cross check (see Table 44).

Table 44: Modelled evolution of AMS scores: Belimumab week 24 responders – Target population

Year	Belimumab week 24 responders		
	SS	Manufacturer AMS	ERG AMS
0	12.7	12.7	12.7
1	4.9	9.9	8.8
2	3.4	8.4	7.0
3	2.6	7.4	5.9
4	2.3	6.8	5.2
5	2.2	6.3	4.7
6	2.2	5.9	4.3
7	2.2	5.6	4.1
8	2.3	5.4	3.9
9	2.4	5.2	3.7
10	2.5	5.0	3.6

The values submitted in response to the ERG clarification question are averages across many individual patient iterations. Some non-linearity or rounding approximation may have crept into the figures. But it is difficult to reconcile the ERG cross check of the AMS for SoC and belimumab week 24 non-responders with the discrepancies between the ERG cross check and the manufacturer reported values of the AMS for belimumab week 24 responders.

It is unclear whether the above discrepancy is due to an error in the manufacturer response to the ERG clarification question, an error in the VB coding of the model or an error in interpretation by the ERG. The manufacturer figures for the AMS may incorporate discontinuations within the belimumab week 24 responder figures.

Ignoring this discrepancy for the moment, by definition the AMS introduces memory of previous SS scores to the modelling. As a consequence of this, even when there is no modelled contemporaneous difference in SS scores between the arms at a particular point, the AMS scores will retain a memory of previous differences in SS scores between the arms. Any errors in the calculation of SS scores within the first year will, even if largely washed out over a relatively short period due to a high discontinuation rate within the belimumab arm, continue to be carried forward by the AMS.

Patients within the JHU cohort were recruited at somewhat lower values in their SS scores compared to the BLISS trials, Figure 6.8 of the MS suggesting an average SS score in their first year of around 2.8 with the subsequent AMS being reasonably level or declining slightly over the 17 years of data presented within Figure 6.8. At a minimum it seems possible that JHU cohort type patients would have a history of lower SS scores prior to being eligible for recruitment to the BLISS trials. Any history of SS scores prior to baseline will tend to dampen the impact that changes in SS scores at baseline have upon the AMS.

As noted by the manufacturer, the AMS tends to smooth out changes in the SS scores. But by not taking into account a patient's SS score prior to the model baseline, the model effectively ignores this smoothing out effect and exaggerates the impact that the changes in SS scores at baseline have upon the AMS and upon the net difference in the AMS between the arms. For the AMS at time T from model baseline for belimumab week 24 responders from the written submission it seems that the intention of the model is to calculate it as:

With the net difference in the AMS at time T between belimumab week 24 responders and SoC as:

But this ignores the previous history of SS scores since diagnosis D years prior to the model baseline. Taking this into account results in an AMS at time T for belimumab week 24 responders of:

With the net difference in the AMS at time T between belimumab week 24 responders and SoC as:

Which implies that:

Manufacturer clarifications on SS scores suggests that SS scores among belimumab week 24 responders are never worse than the contemporaneous SS score for SoC, and typically appear to be modelled as being superior[f]. Since (T+D)/T is greater than one this implies that modelled superiority in AMS for belimumab week 24 responders over SoC systematically overstates the actual superiority in AMS for belimumab week 24 responders over SoC due to not having taken into account the previous patient history.

Any overstatement of effect upon the AMS arising because of this will be larger during the early years of the modelling, and for those patients with a long prior history of SLE. The average duration of disease at recruitment was 6 -7 years within the Target population. The average duration of disease at recruitment to the JHU cohort was a little over 5 years.

Calculation of steroid dose

The ERG expert opinion is that steroid use is variable, but that the tapering allowed within the trials was not unrepresentative of UK practice. In response to ERG clarification question A8 the manufacturer has clarified that within the BLISS trials the evolution of steroid use in the Target population was as shown in Table 45.

Table 45: Average steroid use (mg): BLISS Target population

	BLISS-52		BLISS-76		All BLISS	
	SoC	Belim.	SoC	Belim.	SoC	Belim.
Week 24 responders:	n=63	n=80	n=42	n=50	n=105	n=130
Baseline	12.4 ± 8.6	13.7 ± 10.8	8.9 ± 8.2	10.9 ± 7.6	11.0 ± 8.6	12.6 ± 9.8
Week 24	14.4 ± 11.7	11.5 ± 8.0	9.1 ± 6.7	11.4 ± 8.4	12.3 ± 10.3	11.5 ± 8.1
Week 52	10.5 ± 6.3	8.6 ± 5.9	7.5 ± 6.2	11.9 ± 21.9	9.3 ± 6.4	9.8 ± 14.1
Week 24 non-responders:	n=44	n=32	n=54	n=31	n=98	n=63
Baseline	13.4 ± 8.1	13.6 ± 9.3	11.3 ± 9.2	9.5 ± 8.9	12.2 ± 8.7	11.6 ± 9.3
Week 24	17.8 ± 30.2	14.1 ± 10.1	20.5 ± 54.8	12.5 ± 8.8	19.2 ± 44.8	13.3 ± 9.4
Week 52	13.3 ± 7.0	12.1 ± 9.3	9.9 ± 9.2	9.9 ± 8.3	11.4 ± 8.3	11.0 ± 8.8
Overall:	n=107	n=112	n=96	n=81	n=203	n=193
Baseline	12.8 ± 8.4	13.7 ± 10.4	10.3 ± 8.8	10.4 ± 8.1	11.6 ± 8.6	12.3 ± 9.6
Week 24	15.7 ± 20.6	12.1 ± 8.6	14.9 ± 39.5	11.8 ± 8.5	15.3 ± 30.6	12.0 ± 8.5
Week 52	11.4 ± 6.6	9.4 ± 6.9	8.6 ± 7.8	11.2 ± 18.6	10.2 ± 7.3	10.1 ± 12.9

[f] Note that these are average figures across the 50,000 patient simulated. Within the patient level modelling clinical effectiveness estimates are treated deterministically, and as a consequence this seems likely to apply to each individual belimumab patient and its clone that is modelled within the 50,000 simulations.

Note that the standard deviations associated with these estimates suggests quite strongly skewed data. The manufacturer further notes in response to ERG clarification question A8 that "*corticosteroid taper during the study was determined strictly at the investigators' discretion. There were no protocol mandates regarding dose reduction. The total dose of corticosteroids could be adjusted as clinically required during the first 24 weeks of the study; corticosteroid use beyond pre-specified dose limits resulted in the patient being designated as a non-responder.*"

A crude assessment of the above might suggest that steroid use at baseline is typically lower within BLISS-76 than within BLISS-52, which may reflect the different geographic recruitment for the two trials with BLISS-76 locations being more relevant to the UK.

The modelling ignores the above steroid use data on the grounds of it being unrepresentative of UK practice, choosing instead to use the relationship derived from the JHU cohort. The steroid dose is modelled as a linear function of the AMS score: steroid dose (mg/day) = 3.41 + 0.72 * AMS_T [Table 6.11].

This links with the trial based linear regression of changes in SS scores $SS_{52} = (1+\beta) * SS_0$ which implies that $AMS_{52} \equiv (SS_{52} + SS_0)/2 = (2+\beta)/2 * SS_0$ where for SoC $\beta = -0.349$ and for belimumab week 24 responders $\beta = -0.623$. Given a central baseline of $SS_0 = 12.7$ for the Target population this implies $AMS_{52} = 10.5$ for SoC and $AMS_{52} = 8.7$ for belimumab. For the Target population this in turn implies estimated daily steroid doses at the central baseline SS scores of:

- 12.6mg at baseline
- 11.0mg at week 52 for SoC
- 9.7mg at week 52 for belimumab week 24 responders

While the trial steroid use data appears to be quite skewed, the average steroid doses for the Target population are:

- For BLISS-52
    - 12.8mg at baseline and 11.4mg at week 52 for SoC
    - 13.7mg at baseline and 8.6mg at week 52 for belimumab week 24 responders
- For BLISS-76
    - 10.3mg at baseline and 8.6mg at week 52 for SoC
    - 10.9mg at baseline and 11.9mg at week 52 for belimumab week 24 responders
- For All BLISS

- 11.6mg at baseline and 10.2mg at week 52 for SoC
- 12.6mg at baseline and 9.8mg at week 52 for belimumab week 24 responders

It can be argued that the trial data as presented, and in particular the trial data from BLISS-76, does not triangulate particularly well with the steroid doses estimated within the model[g].

Note also that while the absolute difference in SS scores between a belimumab 24 week responder and her SoC clone is maintained while the patient remains on belimumab, the SoC SS score is modelled as falling over time as shown in Table 6.9 and Figure 6.7 of the MS. As a consequence of this, the absolute difference in steroid dose between a belimumab 24 week responder and her SoC clone will be modelled as falling over time.

Calculation of the CAPD

There is limited detail within the submission on the calculation of the CAPD. It seems possible that this may be subject to the same source of bias as the calculation of the AMS, if a patient's prior history and CAPD to date at baseline is not taken into account.

Calculation of the mortality probability

The JHU model survival excludes age and duration of disease on statistical grounds. The SLE SMRs from the Bernatsky reference[19] are then applied to the ratio of the JHU modelled patient specific probability of death with the JHU modelled SLE average probability of death. This appears quite convoluted. A more natural approach might have been to have reconsidered the treatment of age within the JHU survival model, and given the age ranges within the JHU cohort to assess whether the derived model would be applicable outside a certain age range.

The manufacturer justifies the application of the SLE SMRs as drawn from the Bernatsky reference[19] on the grounds of the JHU survival model being unrepresentative of older patients and so unsuitable for extrapolating into old age. This may be the case, but the SLE SMRs drawn from the Bernatsky reference[19] are larger for younger patients than older patients. Admittedly in the younger age group these will be being applied to lower general population mortality risks.

[g] Within this it should be borne in mind that the week 52 steroid dose for belimumab week 24 responders within the BLISS-76 trial has a particularly high standard deviation.

There is the concern that the multiplication by the Bernatsky SMRs may tend to exaggerate the impact of the covariates within the JHU cohort survival model, and of any differences in the values of the covariates as modelled between the arms of the model.

It can also be noted that the Bernatsky 2006 reference[19] provides a number of cuts of the data for the SMR estimates unadjusted and from a "*multivariate hierarchical regression to determine the independent effects of the factors examined (sex, age group, SLE duration, calendar year period of SLE diagnosis, country) on the relative SMR estimates among SLE patients*". Refer to Table 46.

Table 46: Bernatsky SLE SMRs

Unadjusted SMRs by	Unadjusted SMR	95% CI	Revised SMR	95% CI
Gender				
Female	2.5	2.3 – 2.7	1.2	1.0 – 1.4
Male	1.9	1.7 – 2.2	1.0	(ref)
Age				
16-24	19.2	14.7 – 24.7	..	..
25-40	8.0	7.0 – 9.1	..	..
< 40 (above pooled)	10.7	9.5 – 11.9	6.4	5.5 – 7.5
40-59	3.7	3.3 – 4.0	2.6	2.3 – 3.0
60+	1.4	1.3 – 1.5	1.0	(ref)
Duration SLE years				
< 1	5.4	4.7 – 6.3	7.7	5.9 – 10.2
1 – 4	2.5	2.2 – 2.8	3.2	2.5 – 4.1
5 – 9	2.1	1.9 – 2.4	2.4	1.8 – 3.0
10 – 19	2.0	1.8 – 2.3	1.8	1.4 – 2.2
20+	2.0	1.7 – 2.0	1.0	(ref)

The manufacturer argument for the need to apply the SLE SMRs to the JHU cohort survival model centres on speculation that the JHU cohort survival model does not correctly estimate survival probabilities for older SLE patients who were insufficiently represented within the JHU cohort. A possible approach would be to validate the JHU cohort survival model by examining to what extent the estimates of survival probabilities conform to the SMR estimates given within Bernatsky: both relative to the general population and relative to other SLE patients. If the model results triangulate well with the Bernatsky SMRs it is a good fit, if not it is not a good fit and needs revision. But it would seem sensible to check this first, prior to any ad hoc revisions.

The argument as to why the cohort survival model requires adjustment by the Bernatsky SMRs when modelling patients who are of a similar age to those within the JHU cohort is unclear.

There is also some concern around the SMR values applied from Bernatsky. A recent UK based study found somewhat lower SMRs for SLE as shown in Table 47.

Table 47: SMRs for cohort of UK SLE patients: Caroline Gordon (22 June 2011, personal communication)

Age	SMR	CI
20 - 24	[redacted]	[redacted]
25 – 34	[redacted]	[redacted]
35 – 44	[redacted]	[redacted]
45 – 54	[redacted]	[redacted]
55 – 64	[redacted]	[redacted]
65 – 74	[redacted]	[redacted]
75 – 84	[redacted]	[redacted]
≥ 85	[redacted]	[redacted]

[redacted]

Calculation of the SS score direct effect upon treatment costs

The one year observational cost study divided the data into two six month periods and examined the relationship between the SS score severity class 0, 1, 2 or 3 and the patient's 6 month cost. The SS score severity class was determined by the maximum SS score observed during the relevant 6 month period. To arrive at an annual cost related to the SS score severity class during the six months, the six monthly costs are simply doubled within the submission.

Given that annual costs are required for the model, there is the obvious question of why the cost data is not analysed on an annual basis. SS scores will have varied over the one year observational study, and the maxima are likely to have differed between the two 6 month periods for some if not all patients. It would be anticipated that costs will be highest in the period immediately around any peak in SS score, and will tend to fall away either side of this. Doubling the six monthly cost will tend to have projected a patient's high costs during one period into what was actually a lower cost period.

To labour the point, suppose that a flare leads to an SS score of 12 and a hospital admission of two weeks duration for a particular patient. If the observational cost data had been analysed on a monthly basis and the patient's peak in the SS score due to the flare was observed in the

data, the SS score of 12 or SS score severity class of 3 would in this instance be associated with a hospital admission of two weeks duration. The corollary of the manufacturer approach to annualisation would be to multiply this by 12, leading to the conclusion that in this instance an SS severity index of 3 sees the patient spend half the year admitted to hospital.

While the bias caused by the manufacturer approach will be less than that outlined in the hypothetical example above, it seems likely that it will have tended to exaggerate the association between the SS score average over the year and annual treatment costs. A simpler approach that averages patients' SS scores over the year and relates these to their annual cost would seem to be more in line with the natural history model and probably less likely to lead to bias.

Calculation of HRQoL and Cost impacts of newly incident organ involvement

It can be noted that the assumption that newly incident organ damage will be at the average SLICC score for that SLICC element within the JHU cohort has a possible double impact. As noted by the manufacturer, incident cases will by definition be 1 when incident rather than the average SLICC score for that element within the JHU cohort. But the element being involved at incidence may also tend to be a less serious element. The weighting given to the utility values within each SLICC score is also the average prevalence of the elements within the organ SLICC score within the JHU cohort. If the more serious elements tend to occur later, then not only will the number of elements being involved at incidence be overestimated, their seriousness might be as well.

The above applies with equal force to the costs associated with newly incident organ involvement.

Double counting of treatment costs

Within the manufacturer model there is a direct causal link between the SS score and the incidence of new organ involvement. This is perfectly reasonable. But the costs associated with the SS score are estimated entirely separately from the costs associated with individual organ involvement. Due to the positive association between SS scores and organ involvement, adding the cost associated with SS scores and the cost associated with organ involvement is likely to have double counted these costs to some degree.

The extent of this bias may be limited if rates of organ involvement within the observational cost study conducted during the phase II trial were low. In some sense, there is a need for the corollary of the "clean" utility function on the cost side of the modelling: a "clean" SS score cost function stripped of the impact of organ involvement upon costs.

5.2.9 ERG reconciliation of durations of organ involvement and undiscounted organ costs

The modelling for the target population results in quite large estimates of the net discounted cost savings from reduced pulmonary involvement with belimumab: £3,040 which is around 6% of the total net discounted cost estimate of £51,925. The model output also outlines that the average duration of pulmonary involvement is modelled as 9.87 years within the SoC arm and 9.50 years within the belimumab arm, these appearing to be undiscounted figures (see Table 48).

Table 48: ERG cross check of modelled pulmonary costs – Target population

	SoC	Belimumab	Net
Average survival undiscounted	31.93	34.87	2.93
Pulmonary involvement			
Baseline	2.5%	2.5%	0.0%
Final	39.9%	36.8%	-3.1%
Pulmonary costs cross check			
Pulmonary average duration (D)	9.87	9.50	-0.37
Pulmonary cost year 1 (£Y1)	£9,678	£9,678	
Pulmonary cost year 2+ (£Y2)	£9,603	£9,603	
Total costs (£Y1 + £Y2*(D-1))	£94,896	£91,308	-£3,587
Modelled costs			
Undiscounted costs from model	£94,852	£91,262	-£3,590
Discounted costs from model	£42,692	£39,652	-£3,040
Pulmonary involvement duration			
Pulmonary average duration (affected)	24.75	25.82	1.07

Given this, the modelled average duration of pulmonary involvement can be coupled with the average costs in year 1 and year 2+ for pulmonary involvement to arrive at the average undiscounted costs for pulmonary involvement. This cross check appears to tally very closely with the summary of the model output: within the SoC arm average undiscounted costs for pulmonary involvement of 94,896 compared to the £94,852 reported in the model output, and £91,308 compared to £91,262 for the belimumab arm. This is in part due to year 1 and year 2 costs being very similar for pulmonary involvement which means that the mean organ

duration is sufficient to characterise the undiscounted mean costs; the distribution of organ duration does not have to be taken into account.

But this cross check is based upon the average duration of pulmonary involvement reported within the model output being that applicable across the patient cohort; i.e. including those who are modelled as not developing pulmonary involvement. This implies an average duration of pulmonary involvement among those with pulmonary involvement at baseline or developing pulmonary involvement over the period of the model of 24.75 years for SoC and 25.82 years for belimumab. The has been confirmed as correct by the company that developed the model for the manufacturer, which in turn imples the following average undiscounted durations of organ involvement and average undiscounted organ cost among those having the relevant organ involved at some point during the modelling. Note that organ involvement at baseline was low. These can then be conditioned by the percentages having the relevant organ involved at some point during the modelling to arrive at the avergage organ cost across the cohort as a whole; i.e. including those not having the relevant organ involved at some point during the modelling (see Table 49).

Table 49: Mean undiscounted organ durations and costs – Target population

	Among those with the organ involved				Across the whole patient group			
	Duration		Undiscounted cost		Involvement		Undiscounted cost	
	SoC	Belim	SoC	Belim	SoC	Belim	SoC	Belim
Cardiovascular	23.48	24.53	£14,787	£15,313	24%	21%	£3,527	£3,260
Diabetes	14.72	15.68	£34,408	£36,656	18%	19%	£6,173	£7,035
Gastrointestinal	20.92	22.55	£2,696	£2,697	22%	25%	£595	£675
Malignancy	13.73	14.86	£6,119	£6,120	32%	34%	£1,955	£2,089
Musculoskeletal	23.16	24.83	£40,285	£42,833	49%	49%	£19,552	£20,952
Neuropsychiatric	24.98	26.36	£30,782	£32,349	45%	46%	£13,761	£14,826
Ocular	22.42	23.57	£1,897	£1,917	35%	36%	£666	£690
PV	17.01	18.02	£12,532	£13,130	22%	21%	£2,698	£2,729
Gon. Failure	24.53	25.78	£0	£0	7%	7%	£0	£0
Pulmonary	24.75	25.82	£237,795	£248,049	40%	37%	£94,852	£91,262
Renal	22.16	23.22	£103,220	£108,974	24%	19%	£25,060	£20,947
Skin	31.47	34.11	£0	£0	8%	8%	£0	£0

5.2.10 Comparison with NICE reference case

Table 50 provides a comparison between the MS basecase submission and the NICE reference case.

Table 50: Comparison with NICE reference case

Attribute	Reference case and TA Methods guidance	Does the *de novo* economic evaluation match the reference case
Comparator(s)	Therapies routinely used in the NHS, including technologies regarded as current best practice	The main comparison is between belimumab adjunctive to standard therapy and standard therapy alone The NICE scope also includes rituximab as a comparators
Patient group	As per NICE scope	The manufacturer niches belimumab to those within the anticipated license with an SS score of at least 10
Perspective costs	NHS and Personal Social Services (PSS)	Yes
Perspective benefits	All health effects on individuals	Yes
Form of economic evaluation	Cost-effectiveness analysis	Cost utility analysis
Time horizon	Sufficient to capture differences in costs and outcomes	Yes
Synthesis of evidence on outcomes	Systematic review	As there is no consideration of rituximab or cyclophosamide there is no requirement for a synthesis of the evidence as the comparator is the standard care arm of the trials. The only synthesis of the trial data is the pooling of BLISS-52 with BLISS-76
Outcome measure	Quality adjusted life years (QALYs)	Yes
Health states for QALY	Described using a standardised and validated instrument	Yes. The "clean" utility linked to SS scores is derived from EQ-5D trial data The HRQoL impacts from further organ involvement are drawn from a range of studies within the literature
Benefit valuation	Time-trade off or standard gamble	Yes. The "clean" utility linked to SS scores using EQ-5D trial data applies the standard social tariffs from Dolan REFto arrive at utility values
Source of preference data for valuation of changes in HRQL	Representative sample of the public	Yes. The "clean" utility applies the standard social tariffs from Dolan to arrive at utility values
Discount rate	An annual rate of 3.5% on both costs and health effects	Yes
Equity	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	Yes
Probabilistic modelling	Probabilistic modelling	Probabilistic modelling is presented for the base case results
Sensitivity analysis		A wide range of univariate sensitivity analyses and scenario analyses are included

5.3 ERG additional scenario and sensitivity analysis

The model runs 50,000 patient simulations for reliable convergence which takes some time.

[redacted]

[redacted]

Belimumab administration cost

The £126 per administration for two hours of senior nursing time cross checks with the 2009-10 PSSRU health care costs, though including qualification costs increases this slightly to £140 per administration.

ERG expert opinion is in line with the manufacturer in noting that the administration cost for belimumab would be similar to that for tocilizumab which is also a one hour IV infusion. Tocilizumab was recently reviewed by NICE for rheumatoid arthritis in TA 198 within which there was clearly some debate throughout over the assessment up to the FAD about the appropriate administration cost to apply. The ERG noted the availability of tariffs and reference costs for Health Research Group (HRG) codes HD23A to HD23C for Inflammatory Spine, Joint or Connective Tissue Disorders. Refer to Table 51 and Table 52.

Table 51: 2011 NHS Tariffs

Combined Daycase / Elective tariff		
HD23A	with Major CC	£1,730
HD23B	with CC	£595
HD23C	without CC	£471

Table 52: 2009 - 10 reference costs

Daycase		
HD23A	with Major CC	£ 455
HD23B	with CC	£ 412
HD23C	without CC	£ 432

Within the manufacturer's assessment of these costs, the £432 day case reference cost is adjusted pro rata by the number of hours required to yield an administration cost of £115 (£432 * 7.5/2.0). The manufacturer argues that adopting £126 per administration is as a result conservative. Adjusting the reference cost in this manner may be questionable.

Within TA198 the ERG noted that "*The administration cost of each infusion of tocilizumab was assumed to be £142. This originated from the first version of the Birmingham Rheumatoid Arthritis Model (BRAM) model calculations using 0.5 day case administration cost from the 2001 version of the PSSRU Unit Costs of Health and Social Care. That*

administration cost was calculated to be £124… This has then been inflated from 2004 to 2008 to get £142 which, according to the submission, has since been used in a couple of STAs including the Abatacept appraisal (TA141)." Further correcting this for inflation led to a final administration cost of £154 with the Final Appraisal Determination (FAD) indicating that while there was uncertainty around this cost £154 was acceptable, with this flowing through to the costing template for tocilizumab.

Since the day case cost taken from the PSSRU is relatively dated, there is an argument for adopting a similar approach but taking half of the current reference cost of £432 to yield £216 rather than the 7.5/2 suggested by the manufacturer. The most stringent approach would be to apply the full day case cost of £432. Applying these costs[h] results in the following cost effectiveness estimates. Refer to Table 53.

Table 53: Belimumab administration cost sensitivity analyses – Target population

		Without PAS		With PAS	
	SoC	Belimumab	Net	Belimumab	Net
QALYs	9.81	10.61	0.81	10.61	0.81
Admin cost @ £126	..	£9,059	£9,059	£9,059	£9,059
Total cost					
Admin @ £115	£105,366	£156,500	£51,134	[redacted]	[redacted]
Admin @ £126	£105,366	£157,291	£51,925	[redacted]	[redacted]
Admin @ £154	£105,366	£159,304	£53,938	[redacted]	[redacted]
Admin @ £216	£105,366	£163,761	£58,395	[redacted]	[redacted]
Admin @ £432	£105,366	£179,290	£73,924	[redacted]	[redacted]
ICERs					
Admin @ £115			£63,429		[redacted]
Admin @ £126			£64,410		[redacted]
Admin @ £154			£66,907		[redacted]
Admin @ £216			£72,436		[redacted]
Admin @ £432			£91,699		[redacted]

Patient age at baseline

Holding all other variables constant and setting the patient age at baseline to be 30, 40 and 50 results in cost effectiveness estimates of £65,498 per QALY, £62,695 per QALY and £55,439 per QALY respectively; i.e. for otherwise identical patients the cost effectiveness of belimumab improves as the age at first administration increases.

The manufacturer response to ERG clarification question B2 confirmed that these estimates for age 30 and age 50 are correct. The rationale provided by the manufacturer for this initially counter intuitive result lies in the calculated AMS at baseline for a 30 year old and a 50 year old being the same within the hypothetical example. The AMS has the same additive effect within the $\lambda_p = X'\beta$ for the 30 year old as for the 50 year old, since by assumption all patient

[h] Due to time constraints these estimates are derived by applying a multiplier to the total administration costs estimated within the base case run for the Target population. For instance, for the £432 day case cost the multiplier applied is £432/£126 = 3.43.

variables within X'β are the same for the 30 year old and the 50 year old other than the AMS. This rolls through to the patient hazard of death relative to the JHU cohort "average". It seems that it is at this point that the modelling of the 30 year old and the 50 year old diverge, with the age specific Bernatsky SLE SMRs being applied in conjunction with the age specific general population mortality rate.

In the light of what still appears to be counterintuitive result around baseline age, this may argue for a review of the modelling of mortality and the application of the Bernatsky SLE SMRs to the JHU cohort survival model.

Linear regression of SS_{52}-SS_0 central parameter estimates

The base case uses the pooled trial data to estimate the linear regression of SS_{52}-SS_0. As already outlined, the parameter estimates for this regression differ quite considerably between the two trials. All parameter estimates have p values of less than 0.1%.

From Table 54, for BLISS-52 the overall effect for belimumab week 24 responders is a central parameter estimate of -0.6372, as compared to -0.5980 for BLISS-76. But this has to be read in conjunction with the estimates for SoC which are also larger in BLISS-52 than for BLISS-76. The net difference between the overall effect of belimumab week 24 responders over that of SoC is -0.2743 for BLISS-52 and -0.2639 for BLISS-76 which given their similarity as naturally close to that estimated from the pooled data: -0.2742.

Table 54: Linear regression of SS_{52}-SS_0 central parameter estimates – Target population

	BLISS-52	BLISS-76	Pooled
SS_{52}-SS_0 SoC	-0.3629	-0.3341	-0.3493
SS_{52}-SS_0 belimumab	-0.3746	-0.3153	-0.3435
SS_{52}-SS_0 belimumab week 24 responders	-0.2626	-0.2827	-0.2800

For the estimated net SS change among belimumab week 24 responders over that of SoC to be so similar between the trials is surprising given the absolute mean changes as previously reported and repeated in Table 55. There would appear to be a larger absolute advantage within BLISS-52 compared to that within BLISS-76: 3.4 (-7.5 vs. -4.1) compared to 2.8 (-6.8 vs. -4.0).

Table 55: SS changes at week 52 by week 24 responder status and by trial – Target population

	BLISS-52			BLISS-76			Pooled		
Week 24 responder	Resp	NResp	All	Resp	NResp	All	Resp	NResp	All
Mean change from baseline at week 52: SS_{52}-SS_0									
SoC	-5.4	-2.3	**-4.1**	-6.1	-2.3	**-4.0**	-5.7	-2.3	-4.1
Belimumab	**-7.5**	-3.3	-6.3	**-6.8**	-2.5	-5.2	-7.2	-2.9	-5.8

Applying the estimates from the BLISS-52 and BLISS-76 results in economic estimates (see Table 56).

Table 56: Effect upon economic estimates of SS_{52}-SS_0 source: Target population

		Without PAS		With PAS	
	SoC	Belimumab	Net	Belimumab	Net
BLISS-52 as source for SS_{52}-SS_0 regression					
Total cost	£105,195	£157,102	£51,907	[redacted]	[redacted]
QALYs	9.84	10.64	0.80	10.64	0.80
ICER			£64,950		[redacted]
BLISS-76 as source for SS_{52}-SS_0 regression					
Total cost	£105,518	£157,469	£51,951	[redacted]	[redacted]
QALYs	9.77	10.55	0.78	10.55	0.78
ICER			£66,318		[redacted]

As anticipated, the cost effectiveness estimates are not particularly different between the application of the pooled regression coefficients, the BLISS-52 regression coefficients and the BLISS-76 regression coefficients.

Steroid dose use equation

The trial evidence for a steroid dose effect between the arms may be open to question, particularly within BLISS-76. This effect can be removed through a sensitivity analysis that slightly arbitrarily applies constant steroid dose of 10mg/day for all patients[i], together with

[i] Implemented within the *PSA Inputs* worksheet by setting HY7=0 and HZ7=10 for 10mg and 8 for 8mg, and HY7=0. 6799 and HZ7=3.197 for the post 1 Jan 2000 JH data

another that reduces it to 8mg/day. But note that this not only equalises the steroid dose between the arms but also equalises it between patients of differing SS score severity at baseline which is likely to be unrealistic. Unfortunately the electronic model is not easily amended to permit different steroid dosing based upon the individual patient baseline SS score undifferentiated by arm.

Retaining the differentiation of steroid use by arm, in response to ERG clarification question B21 the manufacturer re-estimates the steroid dose equation with a dummy for data that was pre-2000. The dummy was statistically significant and of the anticipated sign at 1.433, with the regression constant and coefficient for the SS score falling to 0.6799. These values can be used for a third sensitivity analysis. Refer to Table 57.

Table 57: Steroid doses: Target population

		Without PAS		With PAS	
	SoC	Belimumab	Net	Belimumab	Net
Constant steroid dose of 10mg/day					
Total cost	£103,261	£154,453	£51,192	[redacted]	[redacted]
QALYs	9.64	10.38	0.74	10.38	0.74
ICER			£68,766		[redacted]
Constant steroid dose of 8mg/day					
Total cost	£104,816	£156,561	£51,745	[redacted]	[redacted]
QALYs	9.79	10.55	0.76	10.55	0.76
ICER			£68,278		[redacted]
Post 1 January 2000 JHU data regression					
Total cost	£105,692	£157,877	£52,186	[redacted]	[redacted]
QALYs	9.84	10.65	0.81	10.65	0.81
ICER			£64,369		[redacted]

Arbitrarily equalising the steroid dose between the arms of the model to a constant 10mg/day or 8/mg per day does affect the overall patient experience and cost, but the net effect of the 10mg/day and the 8/mg is similar. For both, the net costs show limited change from the base case but the net benefits fall away slightly faster, resulting in reasonably similar cost effectiveness estimates of around £68,500 per QALY without the PAS [redacted]. The revised post 1 January 2000 regression has no practical

impact upon the results of the model. This may suggest that the net outcomes of the model are not particularly driven by the level of steroid dose, but differentiated by arm it has some impact with this mainly affecting the QALY side of the cost effectiveness equation.

Modelling of mortality and application of SMR

The requirement to apply the SMRs drawn from the Bernatsky reference within the modelling remains unclear to the ERG, particularly for when the patient being modelled is within the age range of the JHU cohort. But if it is reasonable to apply SLE SMRs within the mortality modelling, there is an additional concern over whether the SMRs from the Bernatsky reference are representative. [redacted]

Table 58: SMRs for sensitivity analysis

Age	Base case	Sens. analysis
16 – 24	19.2	5.3
25 – 40	8.0	3.7
40 – 59	3.7	2.6
60+	1.4	1.4

The SMRs reported in Table 58 result in the following model outputs (Table 59).

Table 59: Sensitivity analysis around SMRs – Target population

		Without PAS		With PAS	
	SoC	Belimumab	Net	Belimumab	Net
Survival LY - undiscounted	35.00	37.58	2.58	37.58	2.58
Total cost - discounted	£116,657	£168,095	£51,438	[redacted]	[redacted]
QALYs - discounted	10.53	11.26	0.726	11.26	0.73
ICER			£70,860		[redacted]

[redacted]

Given the apparently quite large undiscounted additional survival of 2.93 years within the Target population, the contribution of the various coefficients within the JHU cohort survival function of Table 6.12 can be explored. This is most simply achieved by sequentially setting each of the individual coefficients within Table 6.12 to equal zero with the remainder taking the values within Table 6.12 and running the model[j]. It is recognised that , as with the stepwise elimination of coefficients that led to the JHU cohort survival model, further eliminating coefficients within the JHU cohort survival model would if correctly undertaken change the values of the remaining coefficients.

Ordering results by their impacts upon net undiscounted survival results in the following estimates compared to the 2.93 additional life years of the Target population base case: AMS 0.90 life years; renal 2.59 life years; CAPD 2.68 life years; PVD 2.70 life years; age at diagnosis 3.14 life years; diabetes 3.06 life years; gastrointestinal 2.99 life years; malignancy 2.98 life years; and, infection 2.90 life years. Musculoskeletal and cholesterol have no impact upon the net undiscounted survival. Refer to Table 60.

[j] Implemented within the Nat Hist Model Data worksheet by setting the relevant cell within cells AG11:AG52 equal to zero

Table 60: Removing the AMS coefficient from JHU cohort survival function - Target population

		Without PAS		With PAS	
	SoC	Belimumab	Net	Belimumab	Net
Survival LY - undiscounted	20.20	21.10	0.90	21.10	0.90
Total cost - discounted	£127,598	£174,022	£46,424	[redacted]	[redacted]
QALYs - discounted	11.12	11.55	0.43	11.55	0.43
ICER			£106,912		[redacted]

The data reported in Table 60 underlines the importance of the AMS coefficient within the JHU cohort survival function to the anticipated additional 2.93 life years from belimumab use within the Target population. This should be read in conjunction with the concerns around the calculation of the SS score and the resultant calculation of the AMS score. It also highlights the possible significance of applying the Bernatsky SMRs to the patient mortality hazard as drawn from the JHU cohort survival function.

SLICC organ involvement at baseline

The model through random drawings simulates a range of organ involvements at baseline within the 50,000 patient simulated. The central estimate of cost effectiveness average across these. Given this it is illustrative to explore the scenarios of: no organ involvement at baseline; all organs having a SLICC score of 1 at baseline; and, individual organs having a SLICC score of 1 at baseline with no other organ involvement[k]. The net effects reported below relate to the addition of belimumab to SoC. Refer to Table 61.

[k] Implemented within the *Subgroup BLISS data* worksheet by setting cells P245:P256 equal to cell Q64 and cells Q245:T256 equal to 0, and for any organ involvement at SLICC score 1 setting the relevant cell(s) within cells Q245:Q256 equal to cell Q64 with the corresponding cells within P245:P256 equal to 0

Table 61: SLICC involvement at baseline – Target population

SLICC = 1 involvement	None	All	CV	Diabetes*	GI*	Malign*	MSK*
SoC undiscounted LYs	33.45	1.40	33.00	26.05	28.53	22.04	29.84
Net undiscounted LYs	2.82	0.32	3.05	3.13	3.19	3.35	3.13
Net disc. QALYs	0.84	0.13	0.67	0.88	0.86	0.95	0.75
Net disc. Costs ex PAS	£51,018	£16,067	£51,846	£53,135	£51,130	£49,048	£53,094
ICER ex PAS	£60,486	£122,796	£77,635	£60,240	£59,583	£51,759	£71,048
ICER with PAS	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
SLICC = 1 involvement	NP	Ocular	PV*	GF	Pulm	Renal*	Skin
SoC undiscounted LYs	33.05	33.02	23.11	32,87	33.06	26.97	32.86
Net undiscounted LYs	3.08	2.91	3.05	3.00	2.97	2.89	3.11
Net disc. QALYs	0.63	0.83	0.88	0.84	0.61	0.81	0.82
Net disc. Costs ex PAS	£53,303	£51,624	£50,449	£52,530	£65,233	£58,222	£51,033
ICER ex PAS	£84,963	£62,420	£57,486	£62,206	£107,729	£71,932	£61,875
ICER with PAS	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
* Within the JHU cohort survival function of Table 6.12 of the submission							
CV – cardiovascular; GI – gastrointestinal; Malign – malignancy; MSK – musculoskeletal; NP – neuropsychatric; PV – peripheral vascular; GF – gonodal failure; Pulm – pulmonary							

Table 61 illustrates that of the organs not entering the JHU cohort survival function, assuming their individual involvement at baseline with no other organ involvement at baseline has a similar effect upon the anticipated patient survival as there being no organ involvement at all at baseline: an average survival in the SoC arm of a little over 33 years. Belimumab is anticipated to provide around an additional 3 life years.

The impact upon net QALYs is more marked. But it must be borne in mind that within the model multiple organ involvement only sees the HRQoL multiplier for the worst organ involved being applied. Neuropsychiatric involvement or pulmonary involvement are the worst, with HRQoL multipliers of 0.71 and 0.69 respectively, and their involvement from baseline effectively limits QALY gains to those arising from additional survival.

Cardiovascular and musculoskeletal disease with HRQoL multipliers of 0.76 and 0.79 respectively also have this effect but to a lesser extent, as their involvement at baseline leaves open the possibility of the subsequent development of neuropsychiatric involvement and/or pulmonary involvement.

Of those organs entering the JHU cohort survival function, assuming the individual involvement at baseline of malignany or peripheral vascular has the largest impact upon anticipated survival in the SoC arm, the anticipated additional survival from belimumab remains fairly constant at around 3 life years.

Given the individual impacts of organ involvement and the JHU cohort survival function, it may be slightly surprising for the scenario of all organ systems being involved at baseline to result in an average survival within the SoC arm of only 1.40 years. It is only in this admittedly extreme scenario that the anticipated additional survival from belimumab drops noticeably below 3 life years.

Patients may differ at baseline in terms of their organ involvement. For organs within the JHU cohort survival function this is modelled as affecting their anticipated survival under SoC. But almost regardless of their anticipated survival under SoC, adding belimumab to SoC appears to be modelled as yielding a fairly constant additional 3 years survival. This may again highlight the centrality of the modelling of the impact of belimumab on the SS score, and by implication the AMS score, upon model outcomes.

Pulmonary involvement costs and HRQoL

The costs of pulmonary involvement are based upon 90% of patients requiring average direct drug costs of £1571 per month plus 100% of patients requiring £316 other resource use to give a total monthly cost of £1,730. Within this the direct drug cost if only sildenafil was used would be somewhat less at only £348. In response to an ERG clarification question the manufacturer has run two additional sensitivity analyses: one applying the costs from sildenafil and the other excluding all pulmonary costs. These result in cost effectiveness estimates for the Target population of £66,807 per QALY and £68,182 per QALY respectively.

The HRQoL impacts are mainly sourced from the same HTA monograph that examines pulmonary arterial hypertension, these relating to the pulmonary arterial hypertension functional classes II, III and IV. An additional HRQoL value for functional class I is drawn from the Zisman pulmonary fibrosis paper, but as this is assumed to only apply to 1% of

pulmonary arterial hypertension patients it has no impact upon the calculations. This leads to the modelling applying an HRQoL multiplier for pulmonary arterial hypertension involvement of 0.61 as shown in Table 62.

Table 62: Pulmonary arterial hypertension average HRQoL

PAH functional class	HRQoL	% patients
I	0.73	1%
II	0.67	24%
III	0.60	63%
IV	0.52	12%
Weighted Average	0.61	

As already outlined this can be applied within the overall pulmonary HRQoL calculation as shown in Table 63.

Table 63: HRQoL calculation pulmonary involvement from Table 16.19

	HRQoL	JHU %	Weighted	JHU SLICC	Final
Pulmonary hypertension	0.61	33%	0.20		
Pulmonary fibrosis	0.73	42%	0.31		
Shrinking lung (Chest XRay)	1.00	2%	0.02		
Pleural fibrosis (Chest XRay)	1.00	20%	0.20		
Pulmonary infarction/resection	0.94	4%	0.04		
Average across pulmonary			0.77	1.31	0.70

Sensitivity analyses around this parameter do not appear to have been conducted. To explore its impact upon model outputs it can in effect be removed from the modelling, due to the values being treated as multiplicative by setting it equal to 1.00[1]. This results in a central estimate of £65,812 per QALY suggesting that results are not particularly sensitive to this variable.

[1] Implemented within the *QoL Inputs* worksheet by setting cells AG9:AG58=1

6 DISCUSSION

6.1.1 Clinical Effectiveness

Across many outcomes whilst the pooled data appear promising, the effect size for patients in BLISS-52 favoured belimumab to a much greater extent than those in BLISS-76, this applied for both the whole and the Target (high disease activity) populations. The effect sizes in favour of belimumab for the whole population in BLISS-76 were modest and for the most part showed no significant difference between belimumab and placebo groups. BLISS-76 is likely to be more representative of the proposed patient population in England and Wales.

Drawing on FDA data there appeared to be little difference in effectiveness between 1mg/kg and 10mg/kg dose regimens. The reasons for, and implications of, the differences between trials and a lack of dose response between doses for BLISS-76, are worthy of discussion but on available evidence cannot be resolved. The reason BLISS-76 patients were relatively unresponsive to belimumab is unlikely to be attributable to recruitment of patients with inactive disease because all were auto-immune positive at entry and had a SLEDAI score$\geq$ 6 points.

BLISS-76 patients had longer established disease, had more developed organ damage, were older, and were receiving less steroid dosage than those in BLISS-52, and these differences may have contributed to differing responses to therapy. The most obvious differences between trials were in the geographical distribution of study centres and in the racial make-up of the populations. These might be reflected in differences in response to therapy and in the nature of standard of care practices. Ninety two percent of BLISS-76 study centres, but none of the BLISS-52 centres, were located in North America + West Europe. The LBSL02 phase II RCT (100% of the trial centres located in USA + Canada) preceded the BLISS trials and failed to demonstrate effectiveness of belimumab (primary outcomes: percent mean change in SLEDAI score at week 24, and median time to first flare). However, post hoc analysis of LBSL02 data did identify a subgroup of patients (~ 70% of the total) who responded better and who exhibited auto-immune positive disease at trial entry. This population became the focus of the subsequent Phase III BLISS trials. The failure of the LBSL02 trial to show an effect was attributed to the inclusion of inappropriate patients lacking auto-antibodies at recruitment.

Belimumab is an expensive drug and the proposal is that it should be administered at monthly intervals at a dose of 10mg/kg. Because of some doubt regarding the relative effectiveness of differing dose regimens it is possible that in practice belimumab may be used at lower than

10mg/kg; it seems important that if this should happen that good data on effectiveness of reduced dose regimens should be collected.

Target population and proposed licence population

The focus of the MS was the high disease activity "Target population" which represents a subgroup of the proposed "licence population" (in turn a subpopulation of the pooled BLISS population). The primary end point, which was the percentage of responders at week 52 according to the novel composite SRI outcome measure, was very similar for pooled Target population and pooled "licence population" with respectively 19.8% and 24.8% extra responders for belimumab compared to placebo (belimumab vs. placebo odds ratio = 2.7 for both populations). Furthermore the cost-effectiveness of belimumab in each population was essentially the same (base case ICER £64,410 and £66,170 / QALY respectively). Given these results, there appear small grounds on which to distinguish patients in the Target population from those in the proposed licence population on the basis of either clinical or cost-effectiveness and a SLEDAI score cut-off of 10 points, appears to be an arbitrary criterion that would be difficult to implement in practice. One effect of selecting the Target population in preference to the "licence population" is to considerably reduce the manufacturer's calculation of total budget impact of introducing belimumab across the country. (MS section 7).

Belimumab vs. rituximab

No head-to-head trial comparing belimumab with rituximab has been conducted. The ERG and the manufacturer disagree about the commonality of outcome measures available from belimumab and rituximab trials, but concur that a credible indirect comparison is not feasible on the grounds of large difference between trial populations. The ERG note that the primary outcome measure in the relevant Rituximab trial may be a more stringent test of therapeutic effect than that used in the BLISS trials, and therefore are not convinced by the manufacturer's implication that belimumab is necessarily a more effective drug.

Efficacy of belimumab for different SLE manifestations

In the BLISS trials the most commonly involved SLE manifestations were musculoskeletal (60%), mucocutaneous (59%), hematologic (16%), general (11%), renal (11%) and vasculitis (7%). Direct evidence for a beneficial effect of belimumab on other manifestations, such as pulmonary, renal or central nervous system manifestations, is not available.

6.1.2 Cost Effectiveness

The manufacturer presents a complex and impressive natural history model of the evolution of SLE. The visual basic modelling as far as has been assessed to date by the ERG is sophisticated and appears correct. There appear to be some data input discrepancies between the written submission and the electronic model.

It is also unclear from an economic point of view why the manufacturer niches belimumab to those with a baseline SLEDAI score of at least 10. The cost effectiveness estimates for the anticipated license population and the Target population are very similar. Over time it also seems possible that those within the anticipated license population may fall within the Target population.

The base case estimates for a patient falling within the Target population are that belimumab will:

- Increase undiscounted survival by 2.93 years
- Increase discounted patient benefits by 0.81 QALYs
- Increase discounted costs by £51,925 [redacted]
- Cost £64,410 per QALY [redacted]

There are a number of ERG concerns with the modelling of the submission. If these concerns are justified, addressing them appears more likely to worsen the estimate of the cost effectiveness of belimumab than improve it.

- Assuming that belimumab week 24 non-responders will experience the average SS score within the SoC arm seems likely to have over-estimated the average impact upon SS scores within the belimumab arm. The SS scores drive the analysis and any error in their calculation is likely to have a major impact on results
- Not taking into account a patient's history before entry into the trial may further exaggerate the impact upon the AMS of belimumab compared to SoC
- The steroid use data within the trials has been passed over within the modelling.
- The calculation of the cumulative average steroid dose may be subject to a bias similar to that of the calculation of the AMS
- Maintaining the net gain in SS score for a belimumab week 24 responder compared to the parallel patient in the SoC arm while the belimumab week 24 responder remains on treatment may be optimistic
- There is some lack of clarity around the reasons for patients' discontinuation and the derivation of the 8% annual discontinuation rate among belimumab week 24

responders, and of the reasonableness of extrapolating using this value. A low discontinuation rate worsens the cost effectiveness of belimumab

- The requirement to adjust the JHU cohort survival model by SMRs from the literature is unclear and may have tended to exaggerate the impact of the individual covariates within the JHU cohort survival model
- The analysis of the observational cost data on a six monthly basis in order to relate it to the maximum SS score during that period then doubling it to arrive at the annual relationship appears peculiar given that the observational cost data was collected over a year. It may also lead to bias
- The separate estimation of a cost per organ involved may have double counted costs estimated within the SS score cost function to some degree
- There appear to be some discrepancies in the reported model outputs for the average durations of organ involvement, the annual costs of these and the discounted total costs of these organ involvements. There are as a consequence concerns around the calculation of the cost offsets from reduced organ involvement arising from belimumab

6.2 Implications for research

It is unlikely that an industry sponsored trial will be conducted to compare belimumab with rituximab or other new biological interventions for SLE. The cost of a sufficiently powered study to discriminate between such treatments is likely to be too great for such studies to be undertaken independently of industry sponsorship. In view of the relative expense of belimumab and the lack of clear demonstration of a dose response relationship it is possible that in the real world belimumab may be employed at doses less than 10mg/kg. Useful research could be undertaken to monitor such usage and the 24 week response rates elicited.

Due to the paucity of long-term evidence for the continued benefit of belimumab and its safety, monitoring and surveillance of patients who have been treated with belimumab are therefore necessary. Further investigation is needed in patients excluded in the current BLISS-52 and BLISS-76 trials who had severe lupus nephritis or central nervous system manifestations of the disease. The two trials were limited in the inclusions of black patients, who for example account for approximately 25% of lupus patients in the USA. These patients also tend to have more severe disease than the general lupus population. In an earlier Phase II study of belimumab, black patients did significantly better than non-black patients. In contrast the reported Phase III trials found black patients treated with belimumab performed worse than those given placebo. These discrepancies needed to be considered further.

Although BLyS (B-Lymphocyte stimulator) is raised in SLE, reducing its activity with belimumab in SLE patients appears to have only a very modest effect. In RCTs a large proportion of patients in the belimumab group responded, but the placebo group response indicated that many would have responded irrespective of receiving belimumab. In a Targeted population with higher response rates the effect of belimumab remained relatively modest. On this basis, research should be directed at identifying additional factors that independently, or together with BLyS play a role in the pathology of SLE. Until such factors are identified it is probable that the traditional armamentarium of interventions will remain core for the treatment of most SLE patients.

7 REFERENCE LIST

1. ACR Ad Hoc Committee on Systemic Lupus Erythematosus Response Criteria. The American College of Rheumatology response criteria for systemic lupus erythematosus clinical trials: Measures of overall disease activity. *Arthritis and Rheumatism* 2004;50:3418-26.

2. Isenberg DA, Gordon C. From BILAG to BLIPS--disease activity assessment in lupus past, present and future. *Lupus* 2000;9:651-4.

3. U.S.Food and Drugs Administration. *FDA briefing package.* 2010. URL: http://www.fda.gov/downloads/advisorycommittees/committeesmeetingmaterials/drugs/arthritisdrugsadvisorycommittee/ucm233579.pdf (accessed 23 June 2011).

4. Cella D, Yount S, Sorensen M, Chartash E, Sengupta N, Grober J. Validation of the Functional Assessment of Chronic Illness Therapy Fatigue Scale relative to other instrumentation in patients with rheumatoid arthritis. *J Rheumatol* 2005;32:811-9.

5. U.S.Food and Drugs Administration. *Arthritis Advisory Committee Meeting Briefing Document for the 16 November 2010 Meeting: Belimumab: BENLYSTA® (belimumab): Treatment of Systemic Lupus Erythematosus: BLA 125370.* 13/10/2010. URL: http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/ArthritisDrugsAdvisoryCommittee/UCM233581.pdf (accessed 23 June 2011).

6. Wallace DJ, Stohl W, Furie RA, Lisse JR, McKay JD, Merrill JT *et al*. A phase II, randomized, double-blind, placebo-controlled, dose-ranging study of belimumab in patients with active systemic lupus erythematosus. *Arthritis & Rheumatism* 2009;61:1168-78.

7. Navarra SV, Guzman RM, Gallacher AE, Hall S, Levy RA, Jimenez RE *et al*. Efficacy and safety of belimumab in patients with active systemic lupus erythematosus: a randomised, placebo-controlled, phase 3 trial. *Lancet* 2011;377:721-31.

8. van Dis I, Kromhout D, Geleijnse JM, Boer JM, Verschuren WM. Evaluation of cardiovascular risk predicted by different SCORE equations: the Netherlands as an example. *Eur J Cardiovasc Prev Rehabil* 2010;17:244-9.

9. Swaak AJ, van den Brink HG, Smeenk RJ, Manger K, Kalden JR, Tosi S *et al*. Systemic lupus erythematosus: clinical features in patients with a disease duration of over 10 years, first evaluation. *Rheumatology (Oxford)* 1999;38:953-8.

10. Merrill, J. T., Neuwelt, C. M., Wallace, D. J., Shanahan, J. C., Latinis, K. M., Oates, J. C., Utset, T. O., Gordon, C., Isenberg, D. A., Hsieh, H. J., Zhang, D, and Brunetta, P. G. Efficacy and safety of rituximab in patients with moderately to severely active systemic lupus erythematosus: results from the randomized, double-blind phase II/III study EXPLORER. Conference: 72nd Annual Scientific Meeting of the American College of Rheumatology. Abstract L12. 2008. 2008 Oct 24;

11. Gordon C, Merrill JT, Wallace DJ, Hsieh H, Brunetta P. BILAG versus SELENA-SLEDAI scoring in systemic lupus erythematosus (SLE) patients in EXPLORER. *Arthritis and Rheumatism* 2009;60:281.

12. Scottish Intercollegiate Guideline Network. *Search Filters*. 21/04/2011. URL: http://www.sign.ac.uk/methodology/filters.html (accessed 23 June 2011).

13. Wallace DJ, Stohl W, Furie RA, Lisse JR, McKay JD, Merrill JT *et al*. A phase II, randomized, double-blind, placebo-controlled, dose-ranging study of belimumab in patients with active systemic lupus erythematosus. *Arthritis & Rheumatism* 2009;61:1168-78.

14. Navarra SV, Guzman RM, Gallacher AE, Hall S, Levy RA, Jimenez RE *et al*. Efficacy and safety of belimumab in patients with active systemic lupus erythematosus: a randomised, placebo-controlled, phase 3 trial. *Lancet* 2011;377:721-31.

15. Gladman DD, Urowitz MB, Kagal A, Hallett D. Accurately describing changes in disease activity in Systemic Lupus Erythematosus. *J Rheumatol* 2000;27:377-9.

16. Centre for Reviews and Dissemination. *NHS Economics Evaluation Database Handbook*. 2007. URL: www.york.ac.uk/inst/crd/pdf/nhseed-handb07.pdf (accessed 27 June 2011).

17. Dolan P. Modeling valuations for EuroQol health states. *Med Care* 1997;35:1095-108.

18. Department of Health. *NHS reference costs 2005-06*. 07/12/2006. URL: http://webarchive.nationalarchives.gov.uk/+/www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_062884 (accessed 9 April 2011).

19. Bernatsky S, Boivin JF, Joseph L, Manzi S, Ginzler E, Gladman DD *et al*. Mortality in systemic lupus erythematosus. *Arthritis and Rheumatism* 2006;54:2550-7.

20. Buyon JP, Petri MA, Kim MY, Kalunian KC, Grossman J, Hahn BH *et al*. The effect of combined estrogen and progesterone hormone replacement therapy on disease activity in systemic lupus erythematosus: a randomized trial. *Ann Intern Med* 2005;142:953-62.

21. Petri M, Buyon J, Kim M. Classification and definition of major flares in SLE clinical trials. *Lupus* 1999;8:685-91.

22. Petri M, Kim MY, Kalunian KC, Grossman J, Hahn BH, Sammaritano LR *et al*. Combined oral contraceptives in women with systemic lupus erythematosus. *N Engl J Med* 2005;353:2550-8.

8 APPENDICES

8.1 Appendix 1 SLE Flare index

The SLE Flare Index categorizes SLE flare as “mild or moderate” or “severe” based on 6 variables[20,21,22] (check that this is the correct Petri et al, 2005):

- Change in SELENA SLEDAI score from the most recent assessment to current.
- Change in signs or symptoms of disease activity.
- Change in prednisone dosage.
- Use of new medications for disease activity or hospitalization.
- Change in PGA score.
- Hospitalization for SLE activity (severe flare only).

Applied as follows (*Taken from HGS Briefing Document to the FDA Oct 2010*[5]):

SLE Flare Index

Mild to Moderate Flare	Severe Flare
Change (increase) in SELENA SLEDAI score of 3 points or more (but not more than 12)	Change (increase) in SELENA SLEDAI score to > 12
New/worse: Discoid, photosensitive, profundus, cutaneous vasculitis, bullous lupus Nasopharyngeal ulcers Pericarditis Arthritis Fever (SLE)	New/Worse: CNS-SLE Vasculitis Nephritis Myositis Plts < 60,000 Hemolytic anemia: Hb <70 g/L or decrease in Hb > 30 g/L **Requiring**: double prednisone, or prednisone increase to > 0.5 mg/kg/day, or hospitalization
Increase in prednisone, but not to > 0.5 mg/kg/day	Increase in prednisone to > 0.5 mg/kg/day
Added NSAID or hydroxychloroquine for SLE activity	New cyclophosphamide, azathioprine, methotrexate, or mycophenolate for SLE activity
≥ 1.0 increase in PGA score, but not to more than 2.5	Hospitalization for SLE activity Increase in PGA score to > 2.5.

8.2 Appendix 2 Assessment of manufacturer's search strategies

Appraised using PRESS CHECKLIST

Checklist developed by: Sampson M, McGowan J, Lefebvre C, Moher D, Grimshaw J. PRESS: Peer Review of Electronic Search Strategies. Ottawa: Canadian Agency for Drugs and Technologies in Health; 2008. Available from: http://www.cadth.ca/media/pdf/477_PRESS-Peer-Review-Electronic-Search-Strategies_tr_Appendices.pdf

Search for non-randomised studies

1. Translation: Is the search question translated well into search

concepts?

☑ Adequate

☐ Needs revision Provide an explanation or example

_____

_____

2. Operators: Are there any mistakes in the use of Boolean or

proximity operators?

☑ Adequate

☐ Needs revision Provide an explanation or example

_____

_____

3. Subject headings: Are any important subject headings missing or have any

irrelevant ones been included?

☑ Adequate

Needs revision Provide an explanation or example

_____

_____

4. Natural language: Are any natural language terms or spelling variants missing, or

have any irrelevant ones been included? Is

truncation used optimally?

☑ Adequate

☐ Needs revision Provide an explanation or example

_____

_____

5. Spelling & syntax: Does the search strategy have any spelling mistakes, system syntax errors, or wrong line numbers?

☑ Adequate

☐ Needs revision Provide an explanation or example

_____

_____

6. Limits: Do any of the limits used seem unwarranted or are any potentially helpful limits missing?

☑ Adequate

☐ Needs revision Provide an explanation or example

_____

Uses SIGN's Observational study filters for Embase and Medline)

7. Adapted for db: Has the search strategy been adapted for each database to be searched?

☑ Adequate

☐ Needs revision Provide an explanation or example

_____

_____

Other notes:

Initial number in report is 14. Number in combined total for databases is 14.

Search doesn't include comparators, but section 5.1.1 of report implies that their plan was only to look for non-RCTs for belimumab.

Search for RCTs

1. Translation: Is the search question translated well into search concepts?

☑ Adequate

☐ Needs revision Provide an explanation or example

_____

_____

2. Operators: Are there any mistakes in the use of Boolean or proximity operators?

☑ Adequate

☐ Needs revision Provide an explanation or example

_____

_____

3. Subject headings: Are any important subject headings missing or have any irrelevant ones been included?

☑ Adequate

Needs revision Provide an explanation or example

_____

_____

4. Natural language: Are any natural language terms or spelling variants missing, or have any irrelevant ones been included? Is truncation used optimally?

☑ Adequate

☐ Needs revision Provide an explanation or example

_____

_____

5. Spelling & syntax: Does the search strategy have any spelling mistakes, system

syntax errors, or wrong line numbers?

☑ Adequate

☐ Needs revision Provide an explanation or example

_____

_____

6. Limits: Do any of the limits used seem unwarranted or are

any potentially helpful limits missing?

☑ Adequate

☐ Needs revision Provide an explanation or example

_____

Uses SIGN RCT filter for Embase, Sections of the search for Medline appear exactly the same as the SIGN RCT filter for Medline, but several lines are missing covering relevant publication types and other small differences are noted. SIGN filter may have been updated and the version used here is older? – new publication types)

---

7. Adapted for db: Has the search strategy been adapted for each database to be

searched?

☑ Adequate

☐ Needs revision Provide an explanation or example

_____

_____

Other notes:

Initial number in flow diagram in section 5.1 of report is 3774. Number in combined total for databases in Appendix is 3776

In section 5.1 of the report it is stated that 39 full publications and 4 conference proceedings were included, but no details have been given for most of these (only 11 are mentioned in Tables 5.2, 5.3 and 5.4). Ideally we should see a list of all 43 publications in tabular form with clear reasons for exclusion.

Search for economic studies

1. Translation: Is the search question translated well into search

concepts?

Adequate

☑ Needs revision Provide an explanation or example

The searches are not as well done as the searches for the clinical effectiveness section. In Pubmed, the use of title/abstract in the belimumab section of the search strategy has resulted in 7 fewer hits compared to the same sections in the clinical effectiveness search strategy. In Embase several lines include major mistakes resulting in it being unclear how the database would have performed the search. For example, line #1 starts with "exp AND" and includes two EMTREE headings that do not exist: 'lupus'/exp, 'sle'/exp. Testing this line of the search by entering it exactly as reported results in 23221 (06/05/11), which is far fewer than the 58059 reported in the search strategy. There are similar problems in the economic filter section of the search strategy.

The search strategies do not include comparators, but they state in report that they do not intend to search for these (is this reasonable?).

---

2. Operators: Are there any mistakes in the use of Boolean or

proximity operators?

☑ Adequate

☐ Needs revision Provide an explanation or example

---

3. Subject headings: Are any important subject headings missing or have any

irrelevant ones been included?

☐ Adequate

☑ Needs revision Provide an explanation or example

The basic search in Pubmed for lupus was automatically mapped by Pubmed, resulting in the inclusion of the MeSH heading Lupus Vulgaris. The MeSH heading Lupus Erythematosus, Systemic was not included. However, because Pubmed also searched for lupus in all fields papers with this MeSH heading would have been picked up.

---

4. Natural language: Are any natural language terms or spelling variants missing, or

have any irrelevant ones been included? Is

truncation used optimally?

☑ Adequate

☐ Needs revision Provide an explanation or example

---

5. Spelling & syntax: Does the search strategy have any spelling mistakes, system

syntax errors, or wrong line numbers?

☑ Adequate

☐ Needs revision Provide an explanation or example

---

6. Limits: Do any of the limits used seem unwarranted or are

any potentially helpful limits missing?

Adequate

☑ Needs revision Provide an explanation or example

As the numbers found in the subject part of the search were so small for Medline (44 in Pubmed), the use of a filter was inappropriate. It is stated in section 6.1 of report that the CRD sensitive economics filters for Pubmed and Embase were used, but the version used in Pubmed does not match that given in CRD's NHS Economics Evaluation Database Handbook 2007. The versions of the filter that used to be on CRD's website are no longer there after the website restructure so it is possible that they have been updated. However, there are some clear discrepancies in the translation of some elements (i.e.the MeSH heading in the filter exp "Costs and Cost analysis"/ was entered as costs AND "cost analysis". Fortunately, this was translated by Pubmed correctly, but as well as including the correct MeSH heading several odd combinations of free text terms were searched, such as (costs"[All Fields] AND "cost"[All Fields] AND "analysis"[All Fields]). Many of the lines in Embase that should have been searching the EMTREE headings were entered

very differently and resulted in massively fewer hits (e.g. the EMTREE heading in the filter exp Economic Evaluation/ (which when tested brought back 166263 hits on 06/05/11) was entered as "exp AND economic AND ('evaluation'/exp OR evaluation)" and resulted in only 977 hits.

---

7. Adapted for db: Has the search strategy been adapted for each database to be

searched?

☑ Adequate

☐ Needs revision Provide an explanation or example

_____

_____

Other notes:

The initial number in the flow diagram tallies with those in the search strategies.

In section 5.1 of the report it is stated that the 14 excluded studies are listed in section 9.1, Appendix 10, but only 2 are actually listed in this Appendix.

8.3 Appendix 3 List of 43 publications from manufacturer's clinical study search

Inclusion / exclusion table from manufacturer's clarification document

Table A16.1 Summary of publications of RCTs reviewed and their reasons for exclusion

Publication	Reason for exclusion
1. Wallace DJ, Stohl W, Furie RA, Lisse JR, McKay JD, Merrill JT, *et al*. A phase II, randomized, double-blind, placebo-controlled, dose-ranging study of belimumab in patients with active systemic lupus erythematosus. Arthritis Care and Research. 2009 15;61 (9):1168-78.	Included (LBSL02).
2. Furie RA, Petri MA, Wallace DJ, Ginzler EM, Merrill JT, Stohl W, *et al*. Novel evidence-based systemic lupus erythematosus responder index. Arthritis & Rheumatism. [Research Support, N.I.H., Extramural]. 2009 Sep 15;61(9):1143-51.	Included. Linked to LBSL02.
3. Furie R, Stohl W, Ginzler EM, Becker M, Mishra N, Chatham W, *et al*. Biologic activity and safety of belimumab, a neutralizing anti-B-lymphocyte stimulator (BLyS) monoclonal antibody: a phase I trial in patients with systemic lupus erythematosus. Arthritis research & therapy. 2008;10 (5):R109.	Included (LBSL01).
4. Navarra S, Ilianova E, Bae SC, Guzman R, *et al*. Belimumab, a BLYS-specific inhibitor, reduced disease activity, flares, and steroid use in patients with seropositive systemic lupus erythematosus (SLE): BLISS-52 study. EULAR. 2010:Abstract SAT0204.	Included (BLISS-52).
5. Tanasescu C, Gallacher A, Garcia M, Littlejohn G, Saaibi D, *et al*. Belimumab, a BLYS-specific inhibitor, significantly improved physical functioning, fatigue, and other health-related quality of life (HRQOL) measures in patients with seropositive systemic lupus erythematosus (SLE): BLISS-52 study. EULAR. 2010:abstract SAT0206.	Included (BLISS-52).
6. D'Cruz D, Tanasescu C, Navarra S, Guzman R, *et al*. Belimumab, a BLYS-specific inhibitor, reduced disease activity, flares and prednisone use in patients with active seropositive SLE: Phase 3 BLISS-52 study. BSR. 2010: abstract OP3.	Included (BLISS-52).
7. Furie R, Zamani O, Wallace D, Tegzova D, *et al*. Belimumab, a BLyS-Specific Inhibitor, Reduced Disease Activity and Severe Flares in Seropositive SLE Patients: BLISS-76 Study Results through Wk 76 ACR. 2010:Abstract 1454.	Included (BLISS-76).
8. Petri M, Van Vollenhoven RF, Zamani O, Furie RA, *et al*. Belimumab, a BLyS-Specific Inhibitor, Reduces Disease Activity and Severe Flares in Seropositive Systemic Lupus Erythematosus (SLE) Patients: BLISS-76 Study. International Journal of Rheumatic Diseases; Asia Pacific League of Associations of Rheumatology 2010;13:suppl. 1: 110-5, abstract 0281.	Included (BLISS-76).
9. Wallace DJ, Kalunian KC, Petri MA, Strand CV, *et al*. Epratuzumab Demonstrates Clinically Meaningful Improvements in Patients with Moderate to severe Systemic Lupus Erythematosus (SLE): Results from EMBLEM™, at Phase IIb Study ACR. 2010:Abstract 1452.	Investigational drug. Not yet available in the UK.
10. Carneiro JRM, Sato EI. Randomized double-blind clinical study with methotrexate in systemic lupus erythematosus. (Portuguese]. Revista Brasilliana de Reumamologia. 1999;39 (4):203-10.	No requirement for patients to have active autoantibody-positive systemic lupus erythematosus.
11. Islam N, Hossain M, Atiqul Haq S, Noor Alam M, *et al*. Efficacy and safety of methotrexate (MTX) in articular and cutaneous manifestations of systemic lupus erythematosus. EULAR. 2006:Abstract THU0273.	No requirement for patients to have active autoantibody-positive systemic lupus erythematosus. Focus on articular and cutaneous manifestations only.
12. Merrill JT, Neuwelt CM, Wallace DJ, Shanahan JC, Latinis KM, Oates JC, *et al*. Efficacy and safety of rituximab in moderately-to-severely active systemic lupus erythematosus: The randomized, double-blind, phase II/III systemic lupus erythematosus evaluation of rituximab trial. Arthritis and Rheumatism. 2010 January;62 (1):222-33.	Included.
13. Andrade-Ortega L, Irazoque-Palazuelos F, Lopez-Villanueva R, Barragan-Navarro Y, Bourget-Pietrasanta F, Diaz-Ceballos MDLT, *et al*. Efficacy of rituximab versus cyclophosphamide in lupus patients with severe manifestations. A randomized and multicentre study. [Spanish]. Reumatologia Clinica. 2010 September;6 (5):250-5.	No requirement for patients to have active autoantibody-positive systemic lupus erythematosus. Excluded patients on other immunosuppressants (except antimalarials). Cyclophosphamide is not a relevant comparator.
14. Fortin PR, Abrahamowicz M, Ferland D, Lacaille D, Smith CD, Zummer M. Steroid-sparing effects of methotrexate in systemic lupus erythematosus: A double-blind, randomized, placebo-controlled trial. Arthritis Care and Research. 2008 15;59 (12):1796-804.	No requirement for patients to have active autoantibody-positive systemic lupus erythematosus. Excluded patients taking azathioprine.
15. Carneiro JRM, Sato EI. Double blind, randomized, placebo controlled clinical trial of methotrexate in systemic lupus erythematosus. Journal of Rheumatology. 1999;26 (6):1275-9.	No requirement for patients to have active autoantibody-positive systemic lupus erythematosus.
16. Barile-Fabris L, Ariza-Andraca R, Olguin-Ortega L, Jara LJ, Fraga-Mouret A, Miranda-Limon JM, *et al*. Controlled clinical trial of IV cyclophosphamide versus IV methylprednisolone in severe neurological manifestations in systemic lupus erythematosus. Annals of the Rheumatic Diseases. 2005 Apr;64 (4):620-5.	Included patients with severe neurological involvement. Cyclophosphamide is not a relevant comparator.
17. Fries JF, Sharp GC, McDevitt HO, Holman HR. Cyclophosphamide therapy in systemic lupus erythematosus and polymyositis. Arthritis and Rheumatism. 1973 1973;16 (2):154-62.	No requirement for patients to have active autoantibody-positive systemic lupus erythematosus. Included patients with polymyositis. Cyclophosphamide is not a relevant comparator.

18. Dussán KB, Magder L, Brodsky RA, Jones RJ, Petri M. High dose cyclophosphamide performs better than monthly dose cyclophosphamide in quality of life measures. Lupus. 2008(12):1079-85.	Cyclophosphamide is not a relevant comparator.
19. Gonzalez-Lopez L, Cardona-Munoz EG, Celis A, Garcia-De La Torre I, Orozco-Barocio G, Salazar-Paramo M, *et al*. Therapy with intermittent pulse cyclophosphamide for pulmonary hypertension associated with systemic lupus erythematosus. Lupus. 2004;13 (2):105-12.	Cyclophosphamide is not a relevant comparator. Included patients with CNS lupus and lupus nephritis.
20. Petri M, Brodsky RA, Jones RJ, Gladstone D, Fillius M, Magder LS. High-dose cyclophosphamide versus monthly intravenous cyclophosphamide for systemic lupus erythematosus a prospective randomized trial. Arthritis and Rheumatism. 2010 May;62 (5):1487-93.	Cyclophosphamide is not a relevant comparator. Included patients with CNS lupus and lupus nephritis.
21. Bykerk V, Sampalis J, Esdaile JM, Choquette D, Senecal JL, Danoff D, *et al*. A randomized study of the effect of withdrawing hydroxychloroquine sulfate in systemic lupus erythematosus. New England Journal of Medicine. 1991;324 (3):150-4.	Withdrawal study in patients with stable SLE.
22. Tsakonas E, Joseph L, Esdaile JM, Choquette D, Senecal JL, Cividino A, *et al*. A long-term study of hydroxychloroquine withdrawal on exacerbations in systemic lupus erythematosus. Lupus. 1998;7 (2):80-5.	Withdrawal study in patients with stable SLE.
23. Levy RA, Vilela VS, Cataldo MJ, Ramos RC, Duarte JLMB, Tura BR, *et al*. Hydroxychloroquine (HCQ) in lupus pregnancy: Double-blind and placebo-controlled study. Lupus. 2001;10 (6):401-4.	Study in pregnant patients.
24. Bezerra EL, Vilar MJ, da Trindade Neto PB, Sato EI. Double-blind, randomized, controlled clinical trial of clofazimine compared with chloroquine in patients with systemic lupus erythematosus. Arthritis and rheumatism. 2005(10):3073-8.	No requirement for patients to have active autoantibody-positive systemic lupus erythematosus. Clofazimine not available in the UK. Focus on cutaneous manifestations only.
25. Meinão IM, Sato EI, Andrade LE, Ferraz MB, Atra E. Controlled trial with chloroquine diphosphate in systemic lupus erythematosus. Lupus. 1996(3):237-41.	Chloroquine not available in the UK.
26. Danowski A, Magder L, Petri M. Flares in Lupus: Outcome Assessment Trial (FLOAT), a comparison between oral methylprednisolone and intramuscular triamcinolone. Journal of Rheumatology. 2006 January;33 (1):57-60.	Study in patients presenting with mild or moderate flare.
27. Dougados M, Job-Deslandre C, Amor B, Menkes CJ. Danazol therapy in systemic lupus erythematosus. A one-year prospective controlled trial on 40 female patients. Clinical Trials Journal. 1987;24 (2):191-200.	No requirement for patients to have active autoantibody-positive systemic lupus erythematosus. Danazol is not considered standard of care.
28. Bootsma H, Spronk P, Derksen R, De Boer G, Wolters-Dicke H, Hermans J, *et al*. Prevention of relapses in systemic lupus erythematosus. Lancet. 1995;345 (8965):1595-9.	Study designed to look at prevention of relapses in patients presenting with a rise in anti-dsDNA.
29. Edwards JC, Snaith ML, Isenberg DA. A double blind controlled trial of methylprednisolone infusions in systemic lupus erythematosus using individualised outcome assessment. Annals of the rheumatic diseases. 1987(10):773-6.	Study in patients with severe SLE presenting with an acute exacerbation.
30. Dammacco F, Della Casa Alberighi O, Ferraccioli G, Racanelli V, Casatta L, Bartoli E. Cyclosporine-A plus steroids versus steroids alone in the 12-month treatment of systemic lupus erythematosus. International Journal of Clinical and Laboratory Research. 2000;30 (2):67-73.	No requirement for patients to have active autoantibody-positive systemic lupus erythematosus.
31. Denburg SD, Carbotte RM, Denburg JA. Corticosteroids and neuropsychological functioning in patients with systemic lupus erythematosus. Arthritis and Rheumatism. 1994 Sep;37 (9):1311-20.	Study was designed to assess the effects of corticosteroids on nervous system functioning as well as disease-related symptoms in patients with mild SLE and mild neuropsychiatric symptoms.
32. Hahn BH, Kantor OS, Osterland CK. Azathioprine plus prednisone compared with prednisone alone in the treatment of systemic lupus erythematosus. Report of a prospective controlled trial in 24 patients. Annals of Internal Medicine. 1975 Nov;83(5):597-605.	Study in severe, life-threatening systemic lupus erythematosus.
33. Mackworth-Young CG, David J, Morgan SH, Hughes GR. A double blind, placebo controlled trial of intravenous methylprednisolone in systemic lupus erythematosus. Annals of the rheumatic diseases. 1988(6):496-502.	No requirement for patients to have active autoantibody-positive systemic lupus erythematosus.
34. Tseng CE, Buyon JP, Kim M, Belmont HM, Mackay M, Diamond B, *et al*. The effect of moderate-dose corticosteroids in preventing severe flares in patients with serologically active, but clinically stable, systemic lupus erythematosus: Findings of a prospective, randomized, double-blind, placebo-controlled trial. Arthritis and Rheumatism. 2006 Nov;54 (11):3623-32.	Included patients with inactive disease defined as a S□EDAI score ≤ 4.
35. Mease PJ, Ginzler EM, Gluck OS, Schiff M, Goldman A, Greenwald M, *et al*. Effects of prasterone on bone mineral density in women with systemic lupus erythematosus receiving chronic glucocorticoid therapy. Journal of Rheumatology. 2005 Apr;32 (4):616-21.	Study designed to examine the effects of prasterone on bone mineral density in female patients with mild to moderate systemic lupus erythematosus.
36. Petri MA, Mease PJ, Merrill JT, Lahita RG, Iannini MJ, Yocum DE, *et al*. Effects of prasterone on disease activity and symptoms in women with active systemic lupus erythematosus: Results of a multicentre randomized, double-blind, placebo-controlled trial. Arthritis and Rheumatism. 2004 Sep;50 (9):2858-68.	Included patients with SLEDAI > 2.
37. Petri MA, Lahita RG, Van Vollenhoven RF, Merrill JT, Schiff M, Ginzler EM, *et al*. Effects of prasterone on corticosteroid requirements of women with systemic lupus erythematosus: A double-blind, randomized, placebo-controlled trial. Arthritis and Rheumatism. 2002;46 (7):1820-9.	No requirement for patients to have active autoantibody-positive systemic lupus erythematosus.
38. Sanchez-Guerrero J, Fragoso-Loyo HE, Neuwelt CM, Wallace DJ, Ginzler EM, Sherrer YRS, *et al*. Effects of prasterone on bone mineral density in women with active systemic lupus erythematosus receiving chronic glucocorticoid therapy. Journal of Rheumatology. 2008 August;35 (8):1567-75.	Study designed to examine the effects of prasterone on bone mineral density in female SLE patients.
39. Chang DM, *et al*. Dehydroepiandrosterone treatment of women with mild-to-moderate systemic lupus erythematosus. Arthritis & Rheumatism. 2002;46(11):2924-27.	Included patients with SLEDAI > 2.
40. Hartkamp A, Geenen R, Godaert GLR, Bijl M, Bijlsma JWJ, Derksen RHWM. Effects of dehydroepiandrosterone on fatigue and well-being in women with quiescent systemic lupus	No requirement for patients to have active autoantibody-positive systemic

erythematosus: A randomised controlled trial. Annals of the Rheumatic Diseases. 2010 June;69 (6):1144-7.	lupus erythematosus.
41. Nordmark G, Bengtsson C, Larsson A, Karlsson FA, Sturfelt G, Ronnblom L. Effects of dehydroepiandrosterone supplement on health-related quality of life in glucocorticoid treated female patients with systemic lupus erythematosus. Autoimmunity. 2005 Nov;38 (7):531-40.	No requirement for patients to have active autoantibody-positive systemic lupus erythematosus.
42. Van Vollenhoven RF, Engleman EG, McGuire JL. Dehydroepiandrosterone in systemic lupus erythematosus: Results of a double-blind, placebo-controlled, randomized clinical trial. Arthritis and Rheumatism. 1995 Dec;38 (12):1826-31.	Study in mild to moderate SLE. No requirement for patients to have active autoantibody-positive systemic lupus erythematosus.
43. Gordon C, Wallace DJ, Shinada S, Kalunian KC, Forbess L, Braunstein GD, *et al*. Testosterone patches in the management of patients with mild/moderate systemic lupus erythematosus. Rheumatology. 2008 Mar;47 (3):334-8.	Included patients with mild to moderate SLE defined by SELENA-SLEDAI ≥ 2.

8.4 *Appendix 4 Demographic details for BLISS total and Target populations (MS Table 5.9)*

	BLISS-52				BLISS-76				Pooled Total Population			
	Placebo N = 287	**1mg/kg N = 288**	**10mg/kg N = 290**	**All N = 865**	**Placebo N = 275**	**1mg/kg N = 271**	**10mg/kg N = 273**	**All N = 819**	**Placebo N = 562**	**1mg/kg N = 559**	**10mg/kg N = 563**	**All N = 1684**
SLE Disease duration (yr)[1]												
Mean ± SD	5.93 ± 6.17	4.96 ± 4.58	5.03 ± 5.07	5.31 ± 5.32	7.42 ± 6.72	7.93 ± 7.13	7.20 ± 7.45	7.52 ± 7.10	6.66 ± 6.48	6.40 ± 6.13	6.08 ± 6.42	6.38 ± 6.35
SELENA SLEDAI score												
≥ 10	158 (55.1%)	139 (48.3%)	160 (55.2%)	457 (52.8%)	141 (51.3%)	144 (53.1%)	136 (49.8%)	421 (51.4%)	299 (53.2%)	283 (50.6%)	296 (52.6%)	878 (52.1%)
Mean ± SD	9.70 ± 3.62	9.56 ± 3.78	9.97 ± 3.88	9.75 ± 3.76	9.80 ± 3.97	9.70 ± 3.65	9.52 ± 3.64	9.67 ± 3.□5	9.75 ± 3.79	9.63 ± 3.71	9.75 ± 3.77	9.71 ± 3.76
PGA score												
< 1	43 (15.0%)	38 (13.2%)	32 (11.0%)	113 (13.1%)	33 (12.0%)	39 (14.4%)	51 (18.7%)	123 (15.0%)	76 (13.5%)	77 (13.8%)	83 (14.7%)	236 (14.0%)
1 - < 2	195 (67.9%)	207 (71.9%)	212 (73.1%)	614 (71.0%)	196 (71.3%)	189 (69.7%)	175 (64.1%)	560 (68.4%)	391 (69.6%)	396 (70.8%)	387 (68.7%)	1174 (69.7%)
≥ 2	49 (17.1%)	43 (14.9%)	46 (15.9%)	138 (16.0%)	46 (16.7%)	43 (15.9%)	47 (17.2%)	136 (16.6%)	95 (16.9%)	86 (15.4%)	93 (16.5%)	274 (16.3%)
BILAG organ domain involvement												
at least 1A or 2B	166 (57.8%)	166 (57.6%)	172 (59.3%)	504 (58.3%)	187 □□8.0%)	173 (63.8%)	160 (58.6%)	520 (63.5%)	353 (62.8%)	339 (60.6%)	332 (59.0%)	1024 (60.8%)
at least 1A	52 (18.1%)	58 (20.1%)	54 (18.6%)	164 (19.0%)	37 (13.5%)	38 (14.0%)	24 (8.8%)	99 (12.1%)	89 (15.8%)	96 (17.2%)	78 (13.9%)	263 (15.6%)
SLICC Damage Index score (Mean ± SD)	0.55 ± 0.93	0.60 ± 1.06	0.55 ± 1.00	0.57 ± 1.00	0.99 ± 1.45	1.04 ± 1.39	0.94 ± 1.38	0.99 ± 1.41	0.77 ± 1.23	0.81 ± 1.25	0.74 ± 1.21	0.77 ± 1.23
SLICC Damage Index score = 0	182 (63.4%)	190 (66.0%)	193 (66.6%)	565 (65.3%)	145 (52.7%)	125 (46.1%)	145 (53.1%)	415 (50.7%)	327 (58.2%)	315 (56.4%)	338 (60.0%)	980 (58.2%)
SLICC Damage Index score = 1	70 (24.4%)	56 (19.4%)	60 (20.7%)	186 (21.5%)	66 (24.0%)	76 (28.0%)	62 (22.7%)	204 (24.9%)	136 (24.2%)	132 (23.6%)	122 (21.7%)	390 (23.2%)
SLICC Damage Index score ≥ 2	35 (12.2%)	42 (14.6%)	37 (12.8%)	114 (13.2%)	64 (23.3%)	69 (25.5%)	66 (24.2%)	199 (24.3%)	99 (17.6%)	111 (19.9%)	103 (18.3%)	313 (18.6%)
Proteinuria (g/24 hour)												
≥ 2	21 (7.3%)	26 (9.0%)	19 (6.6%)	66 (7.6%)	11 (4.0□)	7 (2.6%)	15 (5.5%)	33 (4.0%)	32 (5.7%)	33 (5.9%)	34 (6.0%)	99 (5.9%)
Mean ± SD	0.62 ± 1.15	0.63 ± 1.13	0.54 ± 0.91	0.60 ± 1.07	0.39 ± 0.81	0.33 ± 0.65	0.40 ± 0.73	0.37 ± 0.74	0.50 ± 1.00	0.48 ± 0.94	0.48 ± 0.83	0.49 ± 0.93

[1] Time elapsed between date of SLE diagnosis and the date of informed consent.

BLISS whole population serological (MS Tables 5.10)

	BLISS-52				BLISS-76				Pooled Total Population			
	Placebo N = 287	**1mg/kg N = 288**	**10mg/kg N = 290**	**All N = 865**	**Placebo N = 275**	**1mg/kg N = 271**	**10mg/kg N = 273**	**All N = 819**	**Placebo N = 562**	**1mg/kg N = 559**	**10mg/kg N = 563**	**All N = 1684**
Anti-dsDNA positive (≥ 30 IU/mL)	205 (71.4□)	221 (76.7%)	218 (75□2%)	644 (74.5%)	174 (63.3%)	171 (63.1%)	179 (65.6%)	524 (64.0%)	379 (67.4%)	392 (70.1%)	397 (70.5%)	1168 (69.4%)
Anti-Smith positive (≥ 15 U/mL)	101/287 (35.2%)	102/288 (35.4%)	105/287 (36.6%)	308/862 (35.7%)	72/269 (26.8%)	69/269 (25.7%)	75/265 (28.3%)	216/803 □26.9%)	173/556 (31.1%)	171/557 (30.7%)	180/552 (32.6%)	524/1665 (31.5%)
IgG >ULN (16.18 g/L)	146 (50.9%	140 (48.6%)	151 (52.1%)	437 (50.5%)	108 (39.3%)	105 (38.7%)	94 (34.4%)	307 (37.5%)	254 (45.2%)	245 (43.8%)	245 (43.5%)	744 (44.2%)
Complement												
Normal/high C3 and C4	102 (35.5%)	100 (34.7%)	89 (30.7%)	291 (33.6%)	113 (41.1%)	122 (45.0%)	112 (41.0%)	347 (42.4%)	215 (38.3%)	222 (39.7%)	201 (35.7%)	638 (37.9%)
Low C3 or C4, but not both	78 (27.2%)	55 (19.1%)	75 (25.9%)	208 (24.0%)	65 (23.6%)	57 (21.0%)	60 (22.0%)	182 (22.2%)	143 (25.4%)	112 (20.0%)	135 (24.0%)	390 (23.2%)
Low C3 (< 900 mg/L)	132 (46.0%)	148 (51.4%)	147 (50.7%)	427 (49.4%)	116 (42.2%)	100 (36.9%)	115 (42.1%)	331 (40.4%)	248 (44.1%)	248 (44.4%)	262 (46.5%)	758 (45.0%)
Low C4 (< 16 mg/dL)	160 (55.7%)	173 (60.1%)	180 (62.1%)	513 (59.3%)	143 (52.0%)	141 (52.0%)	147 (53.8%)	431 (52.6%)	303 (53.9%)	314 (56.2%)	327 (58.1%)	944 (56.1%)
Low C3 and C4	107 (37.3%)	133 (46.2%)	126 (43.4%)	366 (42.3%)	97 (35.3%)	92 (33.9%)	101 (37.0%)	290 (35.4%)	204 (36.3%)	225 (40.3%)	227 (40.3%)	656 (39.0%)
BLyS (above LOQ, ≥ 0.5 ng/mL)	273/283 (96.5%)	273/285 (95.8%)	281/285 (98.6%)	827/853 (97.0%)	268/271 (98.9%)	267/270 (98.9%)	263/268 (98.1%)	798/809 (98.6%)	541/554 (97.7%)	540/555 (97□3%)	544/553 (98.4%)	1625/1662 (97.8%)

BLISS whole population concomitant medications (MS Tables 5.11)

	BLISS-52				BLISS-76				Pooled Total Population			
	Placebo N = 287	**1mg/kg N = 288**	**10mg/kg N = 290**	**All N = 865**	**Placebo N = 275**	**1mg/kg N = 271**	**10mg/kg N = 273**	**All N = 819**	**Placebo N = 562**	**1mg/kg N = 559**	**10mg/kg N = 563**	**All N = 1684**
Total corticosteroid use	276 (96.2%)	276 (95.8%)	278 (95.9%)	830 (96.0%)	212 (77.1%)	211 (77.9%)	200 (73.3%)	623 (76.1%)	488 (86.8%)	487 (87.1%)	478 (84.9%)	1453 (86.3%)
Prednisone or equivalent												
> 0 to ≤ 7.5 □g/day	84 (29.3%)	72 (25.0%)	74 (25.5%)	230 (26.6%)	86 (31.3%)	81 (29.9%)	80 (29.3%)	247 (30.2%)	170 (30.2%)	153 (27.4%)	154 (27.4%)	477 (28.3%)
> 7.5 to < 2□ mg/day	136 (47.4%)	133 (46.2%)	131 (45.2%)	400 (46.2%)	76 (27.6%)	96 (35.4%)	81 (□9.7%)	253 (30.9%)	212 (37.7%)	229 (41.0%)	212 (37.7%)	653 (38.8%)
≥ 20 mg/day	56 (19.5%)	71 (24.7%)	73 (25.2%)	200 (23.1%)	50 (18.2%)	34 (12.5%)	39 (14.3%)	123 (15.0%)	106 (18.9%)	105 (18.8%)	112 (19.9%)	323 (19.2%)
Antimalarials	201 (70.0%)	195 (67.7%	185 (63.8%)	581 (67.2%)	180 (65.5%)	171 (63.1%)	168 (61.5%)	519 (63.4%)	381 (67.8%)	366 (65.5%)	353 (62.7%)	1100 (65.3%)
Other immunosuppressants	122 (42.5%)	120 (41.7%)	123 (42.4%)	365 (42.2%)	154 (56.0%)	153 (56.5%)	148 (54.2%)	455 (55.6%)	276 (49.1%)	273 (48.8%)	271 (48.1%)	820 (48.7%)
1 immunosupressant	111 (38.7%)	116 (40.3%)	118 (40.7%)	345 (39.8%)	140 (50.9%)	143 (52.8%)	140 (51.3%)	423 (51.6%)	251 (44.7%)	259 (46.3%)	258 (45.8%)	768 (45.6%)
2 immunosupressants	11 (3.8%)	4 (1.4%)	5 (1.7%)	20 (2.3%)	13 (4.7%)	10 (3.7%)	8 (2.9%)	31 (3.8%)	24 (4.3%)	14 (2.5%)	13 (2.3%)	51 (3.0%)
Azathioprine	67 (23.3%)	71 (24.7%)	84 (29.0%)	222 (25.7%)	57 (20.7%)	52 (19.2%)	58 (21.2%)	167 (20.4%)	124 (22.1%)	123 (22.0%)	142 (25.2%)	389 (23.1%)
Methotrexate	35 (12.2%)	24 (8.3%)	20 (6.9%)	79 (9.1%)	60 (21.8%)	53 (19.6%)	39 (14.3%)	152 (18.6%)	95 (16.9%)	77 (13.8%)	59 (10.5%)	231 (13.7%)
Mycophenolate	19 (6.6%)	16 (5.6%)	17 (5.9%)	52 (6.0%)	42 (15.3%)	45 (16.6%)	50 (18.3%)	137 (16.7%)	61 (10.9%)	61 (10.9%)	67 (11.9%)	189 (11.2%)
Cyclosporin	6 (2.1%)	5 (1.7%)	2 (0.7%)	13 (1.5%)	5 (1.8%)	4 (1.5%)	5 (1.8%)	14 (1.7%)	11 (2.0%)	9 (1.6%)	7 (1.2%)	27 (1.6%)
Leflunomide	2 (0.7%)	-	3 (1.0%)	5 (0.6%)	3 (1.1%)	7 (2.6%)	1 (0.4%)	11 (1.3%)	5 (0.9%)	7 (1.3%)	4 (0.7%)	16 (1.0%)
Cyclophosphamide	2 (0.7%)	3 (1.0%)	1 (0.3%)	6 (0.7%)	2 (0.7%)	2 (0.7%)	2 (0.7%)	6 (0.7%)	4 (0.7%)	5 (0.9%)	3 (0.5%)	12 (0.7%)
NSAIDs	59 (20.6%)	56 (19.4%)	58 (20.9%)	173 (20.0%)	119 (43.3%)	114 (42.1%)	101 (37.0%)	334 (40.8%)	178 (31.7%)	170 (30.4%)	159 (28.2%)	507 (30.1%)

Demographic characteristics of Target population (Table A3.1 clarification document)

		BLISS-52		BLISS-76		Combined BLISS	
		SoC	**10mg/kg**	**SoC**	**10mg/kg**	**SoC**	**10mg/kg**
	Total enrolled	**n=107**	**n=112**	**n=96**	**n=81**	**n=203**	**n=193**
Gender	Male	10 (9.3%)	3 (2.7%)	6 (6.3%)	4 (4.9%)	16 (7.9%)	7 (3.6%)
	Female	97 (90.7%)	109 (97.3%)	90 (93.8%)	77 (95.1%)	187 (92.1%)	186 (96.4%)
Race	Caucasian	29 (27.1%)	23 (20.5%)	61 (63.5%)	54 (66.7%)	90 (44.3%)	77 (39.9%)
	Asian	40 (37.4%)	53 (47.3%)	5 (5.2%)	4 (4.9%)	45 (22.2%)	57 (29.5%)
	Black/African American	1 (0.9%)	6 (5.4%)	13 (13.5%)	7 (8.6%)	14 (6.9%)	13 (6.7%)
	Alaskan/Native American	37 (34.6%)	30 (26.8%)	17 (17.7%)	16 (19.8%)	54 (26.6%)	46 (23.8%)
	Hawaiian/Pacific Islander	0	0	0	0	0	0
	Multiracial	1 (0.9%)	0	0	0	1 (0.5%)	0
	Hispanic origin	55 (51.4%)	46 (41.1%)	28 (29.2%)	25 (30.9%)	83 (40.9%)	71 (36.8%)
Age	Years						
	n ≤ 45 yrs	93 (86.9%)	100 (89.3%)	78 (81.3%)	65 (80.2%)	171 (84.2%)	165 (85.5%)
	n > 45 to < 65	14 (13.1%)	12 (10.7%)	17 (17.7%)	16 (19.8%)	31 (15.3%)	28 (14.5%)
	n ≥ 65 to < 75	0	0	1 (10%)	0	1 (0.5%)	0
Weight (kg)	Mean (SD)	62.1 ± 13.9	61.4 ± 14.1	688□± 13.7	70.0 ± 16.7	65.2 ± 14.2	65.0 ± 15.8
	Range	34.7-127.6	39.5-128.5	45.4-108.6	47.0-131.7	34.7-127.6	39.5-131.7
Region & country	USA/Canada	0	0	45 (46.9%)	24 (29.6%)	45 (22.2%)	24 (12.4%)
	W Europe/Israel	0	0	24 (25.0%)	30 (37.0%)	24 (11.8%)	30 (15.5%)
	E Europe	10 (9.3%)	11 (9.8%)	12 (12.5%)	12 (14.8%)	22 (10.8%)	23 (11.9%)
	America excluding USA/Canada	56 (52.3%)	48 (42.9%)	15 (15.6%)	15 (18.5%)	71 (35.0%)	63 (32.6%)
	Asia	39 (36.4%)	53 (47.3%)	0	0	39 (19.2%)	53 (27.5%)
	Australia	2 (1.9%)	0	0	0	2 (1.0%)	0

BLISS Target population disease characteristics (Table A3.1 Clarification document)

		BLISS-52		BLISS-76		Combined BLISS	
		SoC	**10mg/kg**	**SoC**	**10mg/kg**	**SoC**	**10mg/kg**
	Total enrolled	**n=107**	**n=112**	**n=96**	**n=81**	**n=203**	**n=193**
SLE duration yrs; mean (SD)		6.70 ± 6.96	5.26 ± 4.99	7.42 ± 6.40	7.94 ± 7.47	7.04 ± 6.69	6.38 ± 6.28
BILAG organ involvement	At least 1A or 2B	65 (60.7%)	78 (69.6%)	78 (81.3%)	58 (71.6%)	143 (70.4%)	136 (70.5%)
	At least 1A	18 (16.8%)	25 (22.3%)	21 (21.9%)	7 (8.6%)	39 (19.2%)	32 (16.6%)
	At least 1A or 1B	99 (92.5%)	103 (92.0%)	94 (97.9%)	78 (96.3%)	193 (95.1%)	181 (93.8%)
	No A or B	8 (7.5%)	9 (8.0%)	2 (2.1%)	3 (3.7%)	10 (4.9%)	12 (6.2%)
SELENA-SLEDAI mean (SD)		12.6 ± 3.0	12.8 ± 3.6	13.0 ± 3.5	12.4 ± 2.9	12.8 ± 3.3	12.6 ± 3.3
PGA category	0 to 1	15 (14.0%)	13 (11.6%)	8 (8.3%)	8 (9.9%)	23 (11.3%)	21 (10.9%)
	>1 to 2.5	91 (85.0%)	97 (86.6%)	86 (89.6%)	71 (87.7%)	177 (87.2%)	168 (87.0%)
	>2.5 to 3	1 (0.9%)	2 (1.8%)	2 (2.1%)	2 (2.5%)	3 (1.5%)	4 (2.1%)
SLICC Damage index; mean (SD)		0.6 ± 1.0	0.5 ± 0.9	0.8 ± 1.4	0.8 ± 1.2	0.7 ± 1.2	0.6 ± 1.0
Prednisone or equivalent dose	0 mg/day	5 (4.7%)	4 (3.6%)	15 (15.6%)	12 (14.8%)	20 (9.9%)	16 (8.3%)
	>0 - ≤7.5 mg/day	26 (24.3%)	27 (24□□□)	31 (32.3%)	24 (29.6%)	57 (28.1%)	51 (26.4%)
	> 7.5 mg/day	76 (71.0%)	81 (72.3%)	50 (52.1%)	45 (55.6%)	126 (62.1%)	126 (65.3%)
Average prednisone or equivalent dose; mean (SD) mg/day		12.8 ± 8.4	13.7 ± 10.4	□0□3 □□8.8	10.4 ± 8.1	11.6 ± 8.6	12.3 ± 9.6

8.5 Appendix 5 Justification for pooling results across trials

The ERG points for clarification requested "Please provide further justification for pooling results for Target Population". The manufacturer's response is shown in full below:

Pooling studies under nearly identical protocols but with subjects of varying demographic and baseline characteristics can be justified by extension of the same principles outlined in ICH E9 (Statistical Principles of Clinical Trials) for which different centres in a single multicentre trial are pooled together, i.e. adherence to a common protocol that has been implemented in the same way at all centres using the same standardised procedures and evaluation criteria (as has been done in these studies), and a homogeneous treatment effect across studies (as is the case with these studies). In particular, when pooling the data across these studies, we considered study design, inclusion and exclusion criteria relative to disease severity, and whether the studies were run contemporarily such that the SoC treatment options were similar. These studies followed very similar protocols, were of nearly identical design, had identical inclusion and exclusion criteria, and were conducted over the same time period. Nevertheless, given the heterogeneous presentation of SLE disease and the fact that the Phase 3 program was run globally, one should expect to have variation in the patient population, both within the studies (e.g. between different centres) and between the studies (analogous to differences between centres within the same study).

Since it has been established that the conduct of the studies was effectively the same, one must then determine whether the relative treatment effect is different in one study compared with the other study when evaluating whether two studies are similar enough to pool. Each of the Phase 3 studies achieved statistical significance for belimumab 10mg/kg on the pre-specified primary endpoint of SRI response at Week 52; therefore, these nearly identical, studies provide independent replication of results. While pooling is not necessary to establish the effectiveness of belimumab, it was considered appropriate in order to evaluate treatment effects in high disease activity subgroups of interest, given that the individual studies were not designed to provide sufficient power to demonstrate effectiveness within subgroups. When the two Phase 3 studies were pooled a test for a treatment-by-study interaction was undertaken for the SRI analysis and the treatment-by-study interaction was >0.5. Likewise, for the Target Population of high disease activity, the treatment-by-study interaction was >0.7.

Additionally, a multivariate logistic regression model was developed in order to determine predicators of SRI response. Of the characteristics highlighted as being different between the two studies neither age, race, proteinuria, nor raised IgG were predictors of response. SLICC Damage Score and

complement levels (and their interaction terms with treatment) were included in the final model and neither study (p=0.54) nor the treatment-by-study interaction (p=0.95) was a predictor of SRI response. This result further substantiates that the study is not a predictor of SRI response, thus we believe that is reasonable and valid to pool the two studies.