Belimumab: a technological advance for SLE patients? Report of a systematic review and meta-analysis

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2013-002852
Article Type:	Research
Date Submitted by the Author:	02-Apr-2013
Complete List of Authors:	Kandala, Ngianga-Bakwin; University of Warwick, Warwick Medical School Connock, Martin; University of Warwick, Division of Health Sciences, Warwick Medical School Grove, Amy; University of Warwick, Division of Health Sciences, Warwick Medical School Sutcliffe, Paul; University of Warwick, Division of Health Sciences, Warwick Medical School Mohiuddin, Syed; University of Warwick, Division of Health Sciences, Warwick Medical School Hartley, Louise; University of Warwick, Division of Health Sciences, Warwick Medical School Court, Rachel; Warwick University, Division of Health Sciences Cummis, Ewen; McMDCLtd, UK, G12 9TJ, McMaster Development Consultants Gordon, Caroline; University of Birmingham, School of Immunity and Infection, College of Medical and Dental Sciences Clarke, Aileen; University of Warwick, Division of Health Sciences
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Epidemiology, Evidence based practice, Public health
Keywords:	CLINICAL PHARMACOLOGY, EPIDEMIOLOGY, PREVENTIVE MEDICINE, PUBLIC HEALTH

Belimumab: a technological advance for SLE patients? Report of a systematic review and meta-analysis

Ngianga-Bakwin Kandala, Principal Research Fellow[1, 5]
Martin Connock, Senior Research Fellow[1]
Amy Grove, Project Manager[1]
Paul Sutcliffe, Associate Professor[1]
Syed Mohiuddin, Research Fellow[2]
Louise Hartley, Research Associate[1]
Rachel Court, Information Specialist[1]
Ewen Cummins, Health Economist[3]
Caroline Gordon, Professor of Rheumatology and Consultant Rheumatologist[4]
Aileen Clarke, Professor of Public Health & Health Services Research[1]

[1]Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry, UK, CV4 7AL
[2]The University of Manchester, Health Sciences, Manchester, UK, M13 9PL
[3]McMDCLtd, McMaster Development Consultants, Glasgow, UK, G12 9TJ
[4]School of Immunity and Infection, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK, B15 2TT
[5]KEMRI-University of Oxford-Wellcome Trust Collaborative Programme, Malaria Public Health and Epidemiology Group, Centre for Geographic Medicine, Nairobi, Kenya.

Correspondence to:
Ngianga-Bakwin Kandala, PhD
Principal Research Fellow in Health Technology Assessment
Warwick Evidence and Populations Evidence and Technologies
Division of Health Sciences
Warwick Medical School
University of Warwick
Coventry, CV4 7AL
Tel: +44 (0) 24761 50541
Email: N-B.Kandala@warwick.ac.uk

Short Title:

Systematic review on belimumab for SLE

ARTICLE FOCUS

- SLE is a complex multi-organ auto-immune disease subject to relapse and remission.
- Patients almost always have fatigue, skin rashes and arthritis and there is a wide variety of other problems which the disease can cause.
- Belimumab is a new treatment specifically targeted against SLE.

KEY MESSAGES

1. Combining the results from two RCTs suggests that belimumab is clinically effective for SLE patients.
2. However, all outcomes were systematically superior in one trial compared with the other.
3. Different trial conduct and populations mean that it is too early to draw generalisable conclusions.

STRENGTHS AND LIMITATIONS

- At first sight combined meta analytic evidence suggests that belimumab is clinically effective for patients with severe SLE
- We suggest that it is too early to draw strong conclusions because the 2 relevant trials cover different populations in different countries and there may be differences in trial conduct and outcome assessment.

Abstract

Objectives: To undertake a systematic review and meta-analysis to investigate clinical effectiveness of belimumab for patients with SLE and anti-nuclear and/or anti-dsDNA autoantibodies.

Methods: We searched eight electronic databases and reference lists for randomised controlled trials (RCTs) of belimumab against placebo or best supportive care. Quality assessment and random effects meta-analysis were undertaken.

Design: A meta-Analysis of RCTs.

Setting: NA

Participants: 2133 SLE patients

Interventions: NA

Primary and secondary outcome measures: Responder Index (SRI) at week 52.

Results: Three double-blind placebo-controlled RCTs (L02, BLISS-52 BLISS-76) investigated 2133 SLE patients. BLISS-52 and BLISS-76 trials recruited patients with anti-nuclear and/or anti-dsDNA autoantibodies and demonstrated belimumab effectiveness for the SLE Responder Index (SRI) at week 52. Ethnicity and geographical location of participants varied considerably between BLISS trials. Although tests for statistical heterogeneity were negative, BLISS-52 results were systematically more favourable for all measured outcomes. Meta-analysis of pooled 52-week SRI BLISS results showed benefit for belimumab (OR 1.63, 95% CI 1.27-2.09). By week 76, the primary SRI outcome in BLISS-76 was not statistically significant (OR 1.31, 95% CI 0.919-1.855).

Conclusions: Meta-analysis shows a statistically significant benefit of belimumab for patients with SLE and anti-nuclear and/or anti-dsDNA autoantibodies at 52 weeks only. In view of the different populations studied at different locations and the consistently superior results from one trial compared to the other, the generalizability of pooled results should be viewed with caution. Population heterogeneity, geography and/or variation in trial conduct may be hidden confounders. These findings require further replication or explanation before uncritical acceptance of the positive pooled meta-analytic result is accepted.

INTRODUCTION

SLE is an auto-immune disease subject to relapse and remission. Incidence is estimated at between 1.0 and 10.0 per hundred thousand person years using different measures, and prevalence at between 20-70 per 100,000.[1, 2] SLE is a complex multi-organ disease with a number of different manifestations.[3] Patients almost always have fatigue, often have skin rashes and arthritis and there is a wide variety of other problems which the disease can cause.

The American College of Rheumatology has defined 11 classification criteria, including: rash; photosensitivity; oral ulcers; arthritis; serositis; renal or neurological disorder.[4, 5] Assessment of the patient can be difficult, as flares of the disease have to be distinguished from its complications, from comorbidity especially infection, and from adverse effects of medications.[6] SLE is more common in women (in most studies 90% or more of cases are women)[2] and in those from black and ethnic minorities. Recently age-adjusted incidence rates have been produced showing that rates are highest in women aged 40 years and over.[7] Mortality rates show that 5 year survival is high, at over 90%[8, 9] and the overall SMR has been calculated as 2.4.[10]

Antinuclear antibodies are present in virtually all patients with SLE.[11] Anti-ds DNA antibodies are present in 50-60% patients at some point in their disease but often transiently with active disease.[11] The aim of treatment is to maintain normal function whilst suppressing disease activity and preventing organ damage.[6] Achieving these conflicting aims can be difficult. Corticosteroids are the mainstay of treatment. Other drugs used include antimalarials such as hydroxychloroquine, and immunosuppressive drugs such as azathioprine and mycophenolate mofetil. More recently rituximab (a monoclonal antibody which reacts with the CD20 antigen, which is expressed on B cells) has also been used, although the largest trial undertaken to date failed to reach its end point.[12]

Belimumab (Benlysta®) is an IgG1monoclonal antibody which inhibits the activity of the soluble cytokine BLyS (B lymphocyte stimulator; also known as BAFF). In contrast to earlier SLE treatments, belimumab is targeted at the fundamental pathology of SLE and has been widely interpreted as representing a step change in treatment options.

Belimumab was recently licensed in the USA and in Europe for treatment of autoantibody-positive SLE and is the first drug to be so licensed for several decades. The European indication is for severely affected SLE patients with active, autoantibody-positive disease and a high degree of disease activity exemplified by positive anti-dsDNA and low complement despite standard therapy.[13] Belimumab is administered by IV infusion recommended at 10mg belimumab /kg on days 0, 14 and 28, and at 28 day intervals thereafter. A course of belimumab treatment for a 64kg patient using the US list price of $1,477 (£926.37) for a 400 mg vial[14] would be $56,527 (£35,454) per year, and according to the US average whole sale price of $4.432 (£2780)/400mg vial)[15] would be $42,545 (£26,684) per year.

A number of clinical measures have been developed for tracking the progression of SLE[16] and for estimating the effects of treatment.[17] They include the Physician's Global Assessment (PGA); SELENA-SLEDAI (Safety of Estrogen in Lupus National Assessment-Systemic Lupus Erythematosus Disease Activity Index); and the BILAG Index (British Isles Lupus Assessment Group) and the SRI (SLE Response Index). Their major features are summarised in Figure 1. Their complexity means that outside specialised centres they may not be widely used in routine clinical practice.

[Insert Figure 1 here]

Our objective was to synthesise findings from randomised controlled trials of belimumab for patients with SLE and anti-nuclear and /or anti-dsDNA autoantibodies, to make an overall assessment of the performance of this drug in relation to comparator treatments using the SRI and other outcomes (as listed in Figure 1) and to assess the findings of trials in the light of population samples or geographical factors.[18]

METHODS

The study was undertaken as part of work for the National Institute for Health Research, Health technology Assessment programme (Grant funding reference 10/73/01. Further information is available from: www.hta.ac.uk/).

Search scope

We searched for randomised controlled trials investigating belimumab administered i.v. for patients with SLE and anti-nuclear and /or anti-dsDNA autoantibodies. Comparators considered were Belimumab versus placebo and Belimumab versus best supportive care. Outcomes included all disease-related or health-status-related measures. There was no publication year restriction, but the search was restricted to English language references only.

Search strategy

The following eight databases were searched: Cochrane Database of Systematic Reviews; the Cochrane Central Register of Controlled Trials (CENTRAL); DARE; EMBASE; HTA Database; Medline; Pre-Medline; Science Citation Index. Search strategies for these databases used a combination of terms related to the population and interventions listed above; the specific search strategies are provided in Appendix 1. In Medline and EMBASE the subject strategies were combined with search strategies designed to identify randomised controlled trials (See Appendix 1).

Unpublished studies were identified using: Clinical Trials, Current Controlled Trials, EU Clinical Trials Register, UK Clinical Research Network Study Portfolio, National Research Register, WHO Clinical Trials Search Portal, NHS Evidence, Conference Proceedings Citation Index -Science and Google.

In addition, specific websites were searched: Medicines and Healthcare products Regulatory Agency (MHRA); European Medicines Agency (EMEA); US Food and Drug Administration (FDA) and the following specific conference proceedings: American College of Rheumatology; British Society of Rheumatology; European League Against Rheumatism (EULAR).

Inclusion criteria: Two reviewers assessed retrieved publications for inclusion. Publications were included if they described results from RCTs of Belimumab for SLE patients with positive autoantibodies. Any disagreements were resolved with reference to third reviewer.

Date extraction: Potentially relevant publications were obtained in full text and assessed by the same two reviewers. One reviewer extracted data for all specified primary and secondary outcome measures, for adverse events and deaths. A second reviewer checked extracted data.

Quality evaluation: Quality assessment and risk of bias was guided by the CRD checklist[19] based on all information in the included publications which specifies reporting of randomisation, concealment of allocation, group balance, blinding, drop-outs, outcome reporting bias, and whether ITT analysis was used.

Statistical analysis: Unadjusted odds ratios and mean differences for binary and continuous outcomes were calculated respectively. Statistical heterogeneity was calculated using the I^2 statistic.[20] Adjusted outcome measures were tabulated where these were reported. A random effects meta-analysis[21] was undertaken using STATA version 10 software.[22]

RESULTS

Characteristics of included studies

We identified three placebo controlled RCTs of belimumab versus standard care: the phase III trials termed BLISS-52[23] and BLISS-76[24] and a phase II trial (study L02).[25] The PRISMA flow chart shows the process of identification of publications (see Figure 2). We

identified that there is an on-going trial in Asia.[26] All three completed trials appeared to be of good quality; however details of allocation concealment were meagre (Table 1).

[Insert Table 1 here]

[Insert Figure 2 here]

BLISS-52,[23] BLISS-76[24] and study L02[25] have been published in peer reviewed journals, however the fullest accounts in the public domain are in the FDA licensing approval documents[27, 28] and the manufacturer's 2011 submission to NICE.[29] Each of these placebo-controlled randomised trials was designed with multiple randomised groups. In the L02 trial, patients received 1 or 4 or 10 mg/kg of belimumab or placebo, while in the BLISS trials the belimumab dose regimens were 1mg/kg or 10 mg/kg. Both US and European licensing is for the 10mg/kg dose regimen. In this article we focus on efficacy results for the 10mg/kg licensed regimen relative to placebo. We also consider the off licence 1 mg/kg and 4 mg/kg dose regimens for investigation of adverse events.

Centralised, stratified randomisation was reported as used in all three trials and arms were generally well balanced. All three trials recruited predominantly female patients (~90%) and were described as double blind. The two BLISS studies were conducted according to similar protocols.

There were differences in geographical distribution of the study centres and in the resulting ethnic racial make-up of the study populations (Table 2 and Figure 3). Thus in BLISS-76, 70% were Caucasian, 13% native American and 3% Asian respectively, whereas in BLISS-52, 27% were Caucasian, 32% native American and 38% were Asian. Table 3 lists the major outcomes pre-specified in the BLISS trials.

There were additional population differences between BLISS and L02 trials at recruitment. Reporting of results for patients with anti-nuclear and /or anti-dsDNA autoantibodies in L02 was only included for a post-hoc subgroup. Primary outcomes measured in L02 were not comparable with those of the BLISS studies. For these reasons, L02 study results were not included in the meta-analysis of clinical effectiveness. For the BLISS trials a composite primary outcome measure was developed and termed the SLE Response Index (SRI) (see Figure 1 and Table 3). This pre-specified primary end point is the primary outcome investigated in this meta-analysis.

[Insert Table 2 here]

[Insert Figure 3 here]

[Insert Table 3 here]

[Insert Figure 4 here]

Efficacy results for the BLISS trials for major binary outcomes and for the time to first SLE flare are summarised in Figure 4. The pre-specified primary efficacy end point was the proportion of responders at week 52 according to the novel SRI composite outcome. Both trials satisfied this primary end point with a better result for BLISS-52. The difference in percentage responders in the belimumab group relative to placebo group was 14% in BLISS-52, 9.4% in BLISS-76 and 11.8% when pooled across trials using logistic regression[27] and 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 = 0.0207). By week 76, the proportion of SRI responders in the BLISS-76 trial ceased to reach statistical significance; this also held for the odds ratio adjusted by logistic regression (OR1.31, 95% CI: 0.92 – 1.87, p = 0.1323).[28]

For all other binary effectiveness outcomes, and for time to first flare or to first severe flare, the BLISS-52 trial delivered results which were more favourable to Belimumab than did BLISS-76, with the latter results failing to reach a conventional level of statistical significance except for the ≥ 4 point improvement in SLEDAI score at week 52. Results for continuous outcomes are summarised in Figure 5. These revealed a similar pattern of BLISS-52 superiority for all reported outcomes. Mean changes from baseline for FACIT-fatigue scores and for EQ-5D utility scores (belimumab versus placebo) (not pictured) did not reach statistical significance although again, improvement observed in BLISS-52 for these outcomes was superior to that seen in BLISS-76. BLISS-52 showed a systematic superiority over BLISS-76 across the full range of effectiveness outcomes (binary, time to event and continuous).In BLISS-76 the primary outcome response rates were 32% (46 out of 145), and 35% (47 out of 136) for placebo and belimumab respectively for patients from the US and Canada. In comparison, the corresponding rates for patients from Latin America in BLISS-52 were 49% (71 out of 145), and 61% (85 out of 140).

Figure 4 shows the results for major safety outcomes across the three trials (L02 BLISS-52 and BLISS-76). Although there were more serious adverse events, more serious infections and more deaths associated with belimumab than with placebo, none of the odds ratios for these outcomes reached statistical significance. There were 14 deaths during the controlled phase of the three trials; 3 in the placebo group (n=675), and 11 in the belimumab groups (n=1458) with 6 in the 10mg/kg and 5 in the 1mg/kg groups, respectively (odds ratio 11.7; 95% CI 0.474 to 6.124). The causes of death were various.

[Insert Figure 5 here]

Meta-analysis of study level results showed a statistically significant benefit of belimumab for all main outcomes SRI, SELENA SLEDAI, worsening in PGA, BILAG score or steroid use (Figure 6). Tests for statistical heterogeneity were not significant. However in the BLISS-52 study, Physicians' global assessments (PGA) (which also constitute a component of SRI and SELENA SLEDAI) were more positive for change by week 24 by almost 10% than they were for the BLISS-76 study (BLISS-52: placebo 22.44%; belimumab 36.75% and BLISS-76: placebo 26.18%; belimumab 27.57%).

[Insert Figure 6 here]

DISCUSSION

We undertook a systematic review of the clinical effectiveness of Belimumab, a new treatment targeted at systemic lupus erythematosus (SLE) patients with anti-nuclear and /or anti-dsDNA autoantibodies. We performed an extensive search and systematic review of both completed and on-going trials using a number of databases and by checking reference lists. Data were extracted independently and studies were quality assessed. Random effects meta-analysis was undertaken.

We identified three RCTs (L02, BLISS-52 BLISS-76) reporting data on over 2000 patients. In contrast to the BLISS trials, L02 recruited patients were not necessarily current carriers of anti-nuclear or anti dsDNA antibodies at study commencement. L02 failed to demonstrate clinical effectiveness for the primary end points. Meta-analysis of the BLISS studies showed a benefit of belimumab with the main primary outcome (SRI), showing improvement at 52 weeks, (OR 1.63; 95% CI: 1.27-2.09) although by week 76, the proportion of SRI responders in the BLISS-76 trial ceased to reach statistical significance (OR 1.31 (95% CI: 0.92–1.87

p=0.1323). There were no significant differences between placebo and intervention groups in quality of life or adverse events.

We found that the benefits of belimumab were systematically greater across the board (although not significantly so) in the BLISS-52 trial for all outcomes and although tests for statistical heterogeneity were negative, the racial background and ethnicity of participants varied considerably. If the two BLISS trials were drawn from the same underlying populations, whilst one might expect outcomes to differ, they should differ randomly – some better some worse than the other.

A few studies have directly assessed the existence of and importance of geographical differences in trial outcomes.[30-32] Key factors contributing to such differences are variation in underlying patient population characteristics and variation in study execution. Vickers et al[31], found that Eastern Asian and Eastern European studies had a higher proportion of positive trial results when compared to other countries. O'Shea and De Mets also report that within the Beta-Blocker Heart Attack Trial (BHAT), not only was there a difference in the direction, but also in the size of treatment effect between Canada and the US, although it should be noted that the original aim of that trial was not investigation of international differences in treatment effect.[33] One study found that 96-99% of the total variance in the Global utilisation of strategies to open occluded coronary arteries IV acute coronary syndromes (GUSTO IV ACS) trial could be accounted for by patient-level factors.[34]

International trials need to harmonise training of investigators, patient selection, treatment management, thresholds to centre admission, access to facilities, ascertainment of endpoints and, by implication, results of interest.[35-42] and it is possible that in each country's centres these factors may differ systematically.[35]

Equally, underlying differences in populations and countries (ethnicity, genetics, socio-economic status and health-care systems) and the nature and epidemiology of SLE according to ethnic background may result in differences in reporting of outcomes and pooled results.

The outcomes used in the BLISS trials would be unfamiliar to most of the investigators and it is possible that criteria may have differed between countries. In particular the Physician Global Assessment (PGA) is an important element of the outcomes measured (see Figure 1). PGA was measured as an outcome in itself, and it is also incorporated in SRI.

PGA is of particular concern because as a global physician assessment of a patient's SLE status, it is subjective. The investigators reported a nearly 10% difference between the BLISS-52 and BLISS-76 studies in estimates of percentage change in PGA score in intervention groups at week 24 compared to baseline and this single result in one of the two trials is likely to have had an important influence on our findings of the effectiveness of belimumab in SLE patients.

The latest results of belimumap in patients with SLE(phase 2 study design) of 449 patients with active SLE (USA/Canada) show that 177 (39.4%) patients remained on treatment after 7 years of therapy (i.e approximately 1746 cumulative patients-years) with durable sustained improvement in SLE disease activity (SRI and PAG) [43].

CONCLUSIONS

In conclusion, systematic review and random effects meta-analysis of two RCTs of belimumab for patients with autoantibody positive SLE demonstrated positive results in the main outcome at week 52. However in view of the different populations studied at different locations in the BLISS trials and the consistently superior results from one trial compared to the other, we consider that population heterogeneity; geography and/or variation in trial conduct and outcome assessment should be considered as potential hidden confounders. The generalizability of meta-analytically pooled results should be viewed with caution and we suggest that it is too early to draw strong conclusions in this case.

Acknowledgements

The authors would like to thank the National Institute for Health Research, Health Technology Assessment programme for funding this work.

Funding statement

This work was supported by the National Institute for Health Research, Health Technology Assessment programme [grant number 10/73/01].

Competing interest statement

No conflicts of interest.

Contributions:

N-BK : Conception and design. Data analysis and interpretation. Drafting the article. Critical revisions for important intellectual content. Approval of final article for submission.

MC: Conception and design. Data analysis and interpretation. Literature review. Interpretation of results. Drafting the article. Critical revisions for important intellectual content. Approval of final article for submission.

AG: Interpretation of results. Critical revisions for important intellectual content.

PS: Literature review. Interpretation of results. Critical revisions for important intellectual content.

SM: Data analysis and interpretation. Interpretation of results. Critical revisions for important intellectual content.

LH: Literature review. Interpretation of results. Critical revisions for important intellectual content.

RC: Literature review. Critical revisions for important intellectual content.

EC: Interpretation of results. Critical revisions for important intellectual content.

CG: Interpretation of results. Critical revisions for important intellectual content.

AC: Interpretation of results. Critical revisions for important intellectual content.

All authors read and approved the final manuscript.

REFERENCES

1. Bernatsky S, Joseph L, Pineau CA, Tamblyn R, Feldman DE, Clarke AE A population-based assessment of systemic lupus erythematosus incidence and prevalence--results and implications of using administrative data for epidemiological studies. Rheumatology (Oxford) 46; 1814-8.

2. Pons-Estel GJ, Alarcon GS, Scofield L, Reinlib L, Cooper GS Understanding the epidemiology and progression of systemic lupus erythematosus. Semin Arthritis Rheum 39; 257-68.

3. Gordon C, Li CK, Isenberg DA Systemic lupus erythematosus. Medicine 38; 73-80.

4. Hochberg MC Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 40; 1725.

5. Tan EM, Cohen AS, Fries JF *et al.* The 1982 revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 25; 1271-7.

6. Griffiths B, Mosca M, Gordon C Assessment of patients with systemic lupus erythematosus and the use of lupus disease activity indices. Best Practice and Research in Clinical Rheumatology 19; 685-708.

7. Somers EC, Thomas SL, Smeeth L, Schoonen WM, Hall AJ Incidence of systemic lupus erythematosus in the United Kingdom, 1990-1999. Arthritis Rheum 57; 612-8.

8. Pons-Estel GJ, Alarcon GS, McGwin G, Jr. *et al.* Protective effect of hydroxychloroquine on renal damage in patients with lupus nephritis: LXV, data from a multiethnic US cohort. Arthritis Rheum 61; 830-9.

9. Pons-Estel GJ, Gonzalez LA, Zhang J *et al.* Predictors of cardiovascular damage in patients with systemic lupus erythematosus: data from LUMINA (LXVIII), a multiethnic US cohort. Rheumatology (Oxford) 48; 817-22.

10. Bernatsky S, Boivin JF, Joseph L *et al.* Mortality in systemic lupus erythematosus. Arthritis and Rheumatism 54; 2550-7.

11. Kurien BT, Scofield RH Autoantibody determination in the diagnosis of systemic lupus erythematosus. Scand J Immunol 64; 227-35.

12. Merrill JT, Neuwelt CM, Wallace DJ *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 62; 222-33.

13. Summary of product characteristics. 2012. URL: http://www.medicines.org.uk/EMC/medicine/24769/SPC/Benlysta+120+mg+and+400+mg+powder+for+concentrate+for+solution+for+infusion/ Accessed August 2012

14. American College of Rheumatology. Belimumab for Systemic Lupus Erythematosus. 11-3-2011. http://www.rheumatology.org/publications/hotline/2011_03_15_belimumab.asp Accessed January 2012

15. New Drug Bulletin. Belimumab (Benlysta - Human Genome Sciences, Inc). 2012. URL: http://healthcare.utah.edu/pharmacy/bulletins/NDB_228.pdf Accessed January 2012

16. Isenberg DA, Rahman A, Allen E *et al.* BILAG 2004. Development and initial validation of an updated version of the British Isles Lupus Assessment Group's disease activity index for patients with systemic lupus erythematosus. Rheumatology 44; 902-6.

17. Gordon C, Bertsias G, Ioannidis JPA *et al.* EULAR points to consider for conducting clinical trials in Systemic lupus erythematosus. Ann Rheum Dis 68; 470-6.

18. Burgos PI, McGwin G, Jr., Pons-Estel GJ, Reveille JD, Alarcon GS, Vila LM US patients of Hispanic and African ancestry develop lupus nephritis early in the disease course: data from LUMINA, a multiethnic US cohort (LUMINA LXXIV). Ann Rheum Dis 70; 393-4.

19. Khan KS, Ter Riet G, Glanville J, Snowdon AJ, Kleijnen J . Reviews of Research on Effectiveness. CRD's Guidance for Carrying Out or Commissioning Reviews. 2 nd Edition. (ISBN 1900640201) CRD Report No. 4. York: NHS Centre for Reviews and Dissemination (CRD), University of York. 2001. URL: http://www.york.ac.uk/inst/crd/report4.htm; Accessed January 2012

20. Higgins JPT, Green S . Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. URL: www.cochrane-handbook.org Accessed January 2012

21. DerSimonian R, Laird N Meta-analysis in clinical trials. Control Clin Trials 7; 177-88.

22. STATA Statistical analysis and software. 2011. URL: http://www.stata.com/ Accessed April 2012

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

24. Furie R, Petri M, Zamani O *et al.* A phase III, randomized, placebo-controlled study of belimumab, a monoclonal antibody that inhibits B lymphocyte stimulator, in patients with systemic lupus erythematosus. Arthritis Rheum 63; 3918-30.

25. Wallace DJ, Stohl W, Furie RA *et al.* A phase II, randomized, double-blind, placebo-controlled, dose-ranging study of belimumab in patients with active systemic lupus erythematosus. Arthritis & Rheumatism 61; 1168-78.

26. GlaxoSmithKline. GSK1550188 A 52 week study of belimumab versus placebo in the treatment of subjects with SLE located in North East Asia. 2011. URL: http://clinicaltrials.gov/ct2/show/NCT01345253 Accessed January 2012

27. 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 January 2012

28. U.S. Food and Drugs Administration. FDA briefing package. 2010. URL: http://www.fda.gov/downloads/advisorycommittees/committeesmeetingmaterials/drugs/arthritisdrugsadvisorycommittee/ucm233579.pdf Accessed January 2012

29. GlaxoSmithKline. Single Technology Appraisal (STA): Belimumab for the treatment of active autoantibody-positive systemic lupus erythematosus. 2011. URL: http://www.guidance.nice.org.uk/TA/live/13307/56490/56490.pdf Accessed January 2012

30. Pan Z, Trikalinos TA, Kavvoura FK, Lau J, Ioannidis JP Local literature bias in genetic epidemiology: an empirical evaluation of the Chinese literature. PLoS Med 2; e334.

31. Vickers A, Goyal N, Harland R, Rees R Do certain countries produce only positive results? A systematic review of controlled trials. Control Clin Trials 19; 159-66.

32. Zhang D, Freemantle N, Cheng KK Are randomized trials conducted in China or India biased? A comparative empirical analysis. J Clin Epidemiol 64; 90-5.

33. O'Shea JC, DeMets DL Statistical issues relating to international differences in clinical trials. Am Heart J 142; 21-8.

34. Chang WC, Midodzi WK, Westerhout CM *et al.* Are international differences in the outcomes of acute coronary syndromes apparent or real? A multilevel analysis. J Epidemiol Community Health 59; 427-33.

35. White HD International differences: selection, noise, or real? Eur Heart J 21; 339-42.

36. Fox KA, Goodman S, Bigonzi F, Le L, V, Cohen M Inter-regional differences and outcome in unstable angina; analysis of the international ESSENCE trial. Efficacy and Safety of Subcutaneous Enoxaparin in Non-Q-wave Coronary Events. Eur Heart J 21; 1433-9.

37. Sinnaeve PR, Van de Werf FJ Global patterns of health care for acute coronary syndromes. Curr Opin Cardiol 19; 625-30.

38. Faergeman O, Kjekshus J, Cook T *et al.* Differences in the treatment of coronary heart disease between countries as revealed in the Scandinavian Simvastatin Survival Study (4S). Eur Heart J 19; 1531-7.

39. Berger JS, Stebbins A, Granger CB *et al.* Initial aspirin dose and outcome among ST-elevation myocardial infarction patients treated with fibrinolytic therapy. Circulation 117; 192-9.

40. Welsh RC, Chang W, Goldstein P *et al.* Time to treatment and the impact of a physician on prehospital management of acute ST elevation myocardial infarction: insights from the ASSENT-3 PLUS trial. Heart 91; 1400-6.

41. Yusuf S, Flather M, Pogue J *et al.* Variations between countries in invasive cardiac procedures and outcomes in patients with suspected unstable angina or myocardial infarction without initial ST elevation. OASIS (Organisation to Assess Strategies for Ischaemic Syndromes) Registry Investigators. Lancet 352; 507-14.

42. Bleyer AJ, Hylander B, Sudo H *et al.* An international study of patient compliance with hemodialysis. JAMA 281; 1211-3.

43. Merill JT, Furie RA, Wallace DJ, Stohl, Chatham WW, Weinstein A, McKay JD, Ginzler EM, Zhong ZJ, Freimuth WW, Petri MA; for the LBSLO2/99 Study group, ACR, Washington, DC, November 14, 2012.

Appendix 1

Search Strategies

Cochrane Central Register of Controlled Trials (CENTRAL)

CENTRAL searched via Cochrane Library Interface on 18/05/11

1	MeSH descriptor Lupus Erythematosus, Systemic explode all trees	418
2	(lupus NEAR/3 erythematosus) or (systemic* NEAR/3 lupus) or (SLE)	630
3	(#1 OR #2)	703
4	belimumab OR benlysta	6
5	(#3 AND #4)	4

Medline

Medline searched via Ovid Interface on 19/05/11

1	exp Lupus Erythematosus, Systemic/	42025
2	(lupus adj3 erythematosus).tw.	35497
3	(systemic* adj3 lupus).tw.	31639
4	1 or 2 or 3	50358
5	belimumab.mp.	68
6	benlysta.mp.	3
7	5 or 6	68
8	4 and 7	48
9	randomized controlled trial.pt.	305892
10	controlled clinical trial.pt.	82328
11	randomized.ab.	212836
12	placebo.ab.	124063
13	clinical trials as topic.sh.	153987
14	randomly.ab.	154440
15	trial.ti.	91188
16	9 or 10 or 11 or 12 or 13 or 14 or 15	711420
17	exp animals/ not humans.sh.	3582822
18	16 not 17	656689
19	8 and 18	24

RCT search filter used: Cochrane Highly Sensitive Search Strategy for identifying randomized trials in MEDLINE: sensitivity- and precision-maximizing version (2008 revision); Ovid format. Box 6.4.b in the Cochrane handbook. Reference: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org

Medline In-process

Medline In-Process searched via Ovid Interface on 19/05/11

1	exp Lupus Erythematosus, Systemic/	0
2	(lupus adj3 erythematosus).tw.	1213
3	(systemic* adj3 lupus).tw.	873
4	1 or 2 or 3	1236
5	belimumab.mp.	8
6	benlysta.mp.	4
7	5 or 6	10
8	4 and 7	6

Embase

1	belimumab.mp.orexpbelimumab/	427
2	benlysta.mp.	24
3	1 or 2	428
4	exp systemic lupus erythematosus/	50906
5	(lupus adj3 erythematosus).tw.	40637
6	(systemic: adj3 lupus).tw.	36554
7	4 or 5 or 6	59739
8	3 and 7	302
9	random:.tw.	632763
10	placebo:.mp.	250140
11	double-blind:.tw.	116148
12	9 or 10 or 11	796900
13	8 and 12	144

RCT search filter used: Wong, et al. (2006) Best optimization of sensitivity and specificity. Reference: Wong SS, Wilczynski NL, Haynes RB. Developing optimal search strategies for detecting clinically sound treatment studies in EMBASE. J Med Libr Assoc. 2006 Jan;94(1):41-7. PubMed PMID: 16404468; PubMed Central PMCID: PMC1324770.

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1-2
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	3
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	3
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	3-4
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	4
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Appendix 1
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	4
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	4
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	4
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	4
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	5
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	5

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	4
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	NA
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	5
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	Table 2 and Figure 3
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	Table 1
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Figure 1 and Table 3
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	See Figure 6
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	Table 1
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	NA
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	11
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	8-11
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	10
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	11

PRISMA 2009 Checklist

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: **www.prisma-statement.org**.

SELENA-SLEDAI: encompasses 24 weighted items scored dichotomously as present or absent in the previous 10 days, thus improvement or worsening of a manifestation is not captured. Overall disease activity is scored over a range of 0 to 105 points. A minimum clinically meaningful score change = a decrease of 6 points (overall improvement) or an increase of 8 points (overall worsening). A designated change in score (≥ 4 points) between baseline and follow up can be used to dichotomise patients into responders or non-responders for overall disease.

BILAG[16]: Includes 86 items grouped in 8 organ systems to assesses organ system involvement over the last 4 weeks compared to preceding 4 weeks based on physicians intention to treat using classifications ranging from A to E as follows: A = worsening usually requiring intensification of steroids or immunosuppressant treatments; B = worsening usually requiring antimalarials, nonsteroidal anti-inflammatory drugs (NSAIDs), or low dose steroids; C = stable disease (symptomatic therapy); D = improvement; E = system never involved. Unlike SELENA-SLEDAI it can detect worsening or improvement in individual organ system involvement.

PGA: employed to monitor change in patient overall disease activity; typically a visual analogue scale is used ranging between no disease = 0, mild disease = 1, moderate disease = 2, and severe disease = 3.

SRI: A composite instrument (combining elements of SELENA-SLEDAI, BILAG and PGA) developed by belimumab-trialists in conjunction with the US FDA. It allows patients to be dichotomised into responders or non-responders according to predefined assessment criteria in each of the component elements, such as: a SELEN-SLEDAI improvement of ≥ 4 points, plus no worsening in PGA score by > 0.3 points, plus no new BILAG organ system involvement scoring category A in one system or category B in two or more systems. An advantage of SRI, over any one of its components used alone, may be that it can detect SLE improvement in some initial manifestation(s) while guarding against the possibility that worsening in organ systems or overall disease activity might be masked.

PRISMA 2009 Checklist

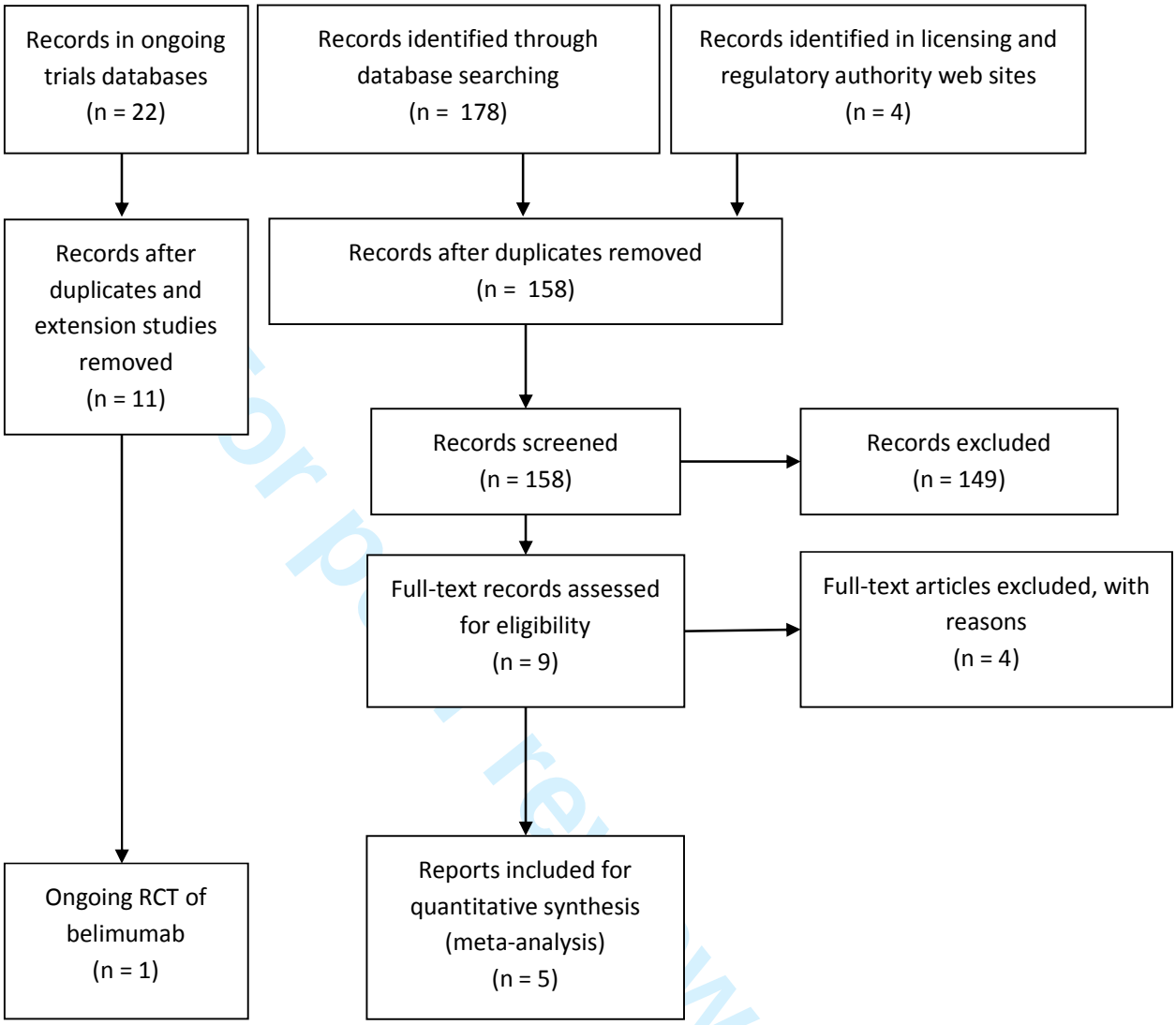
Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	Page 1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	Page 2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	Page 3-4
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	Page 3-4
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	Exists, available from authors
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	Page 5-6
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	Page 5-6
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Appendix
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	Page 5-6
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	Page 5-6
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	Page 5-6
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	Page 8-10
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	Page 8-12
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	Page 5-6

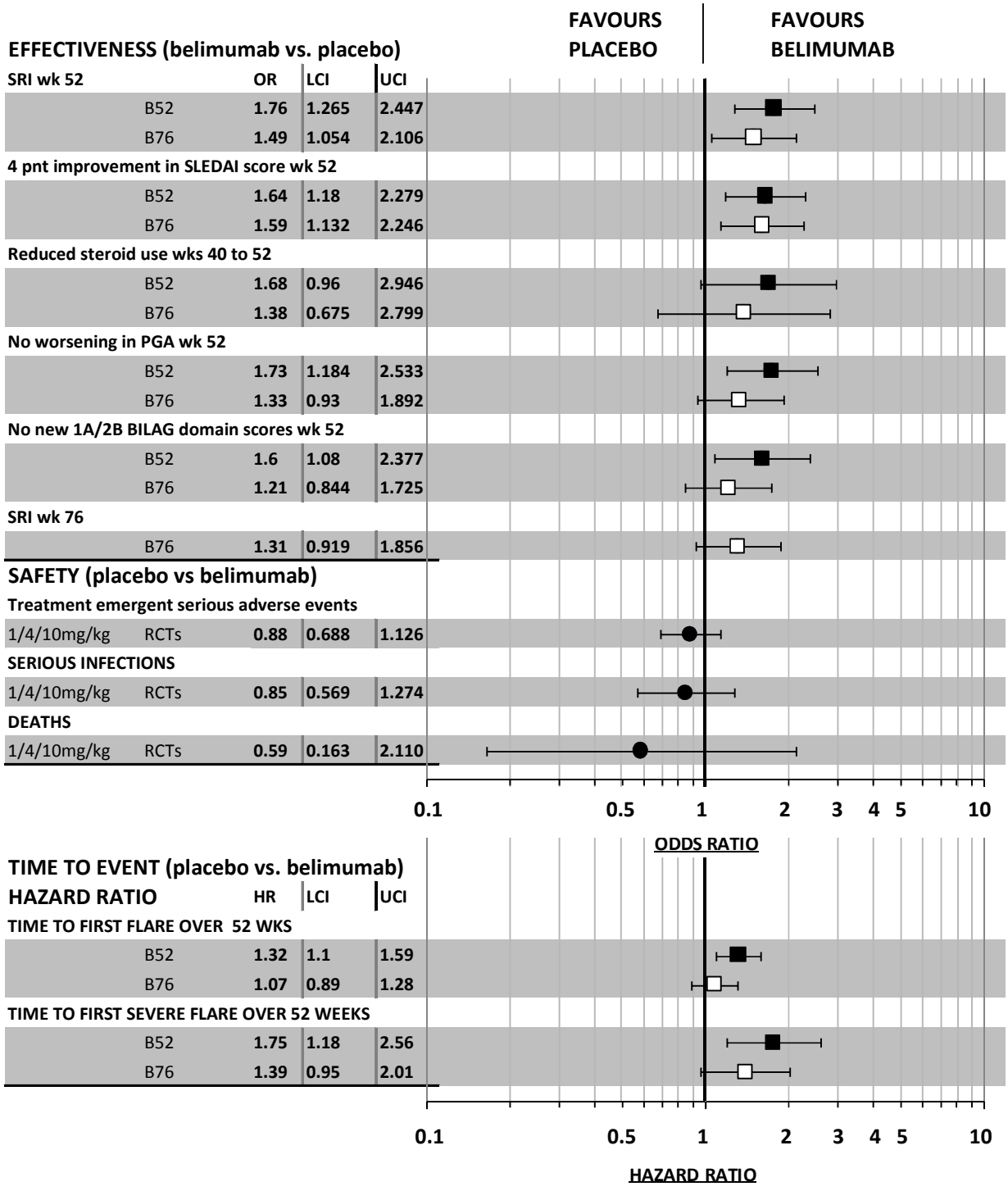
PRISMA 2009 Checklist

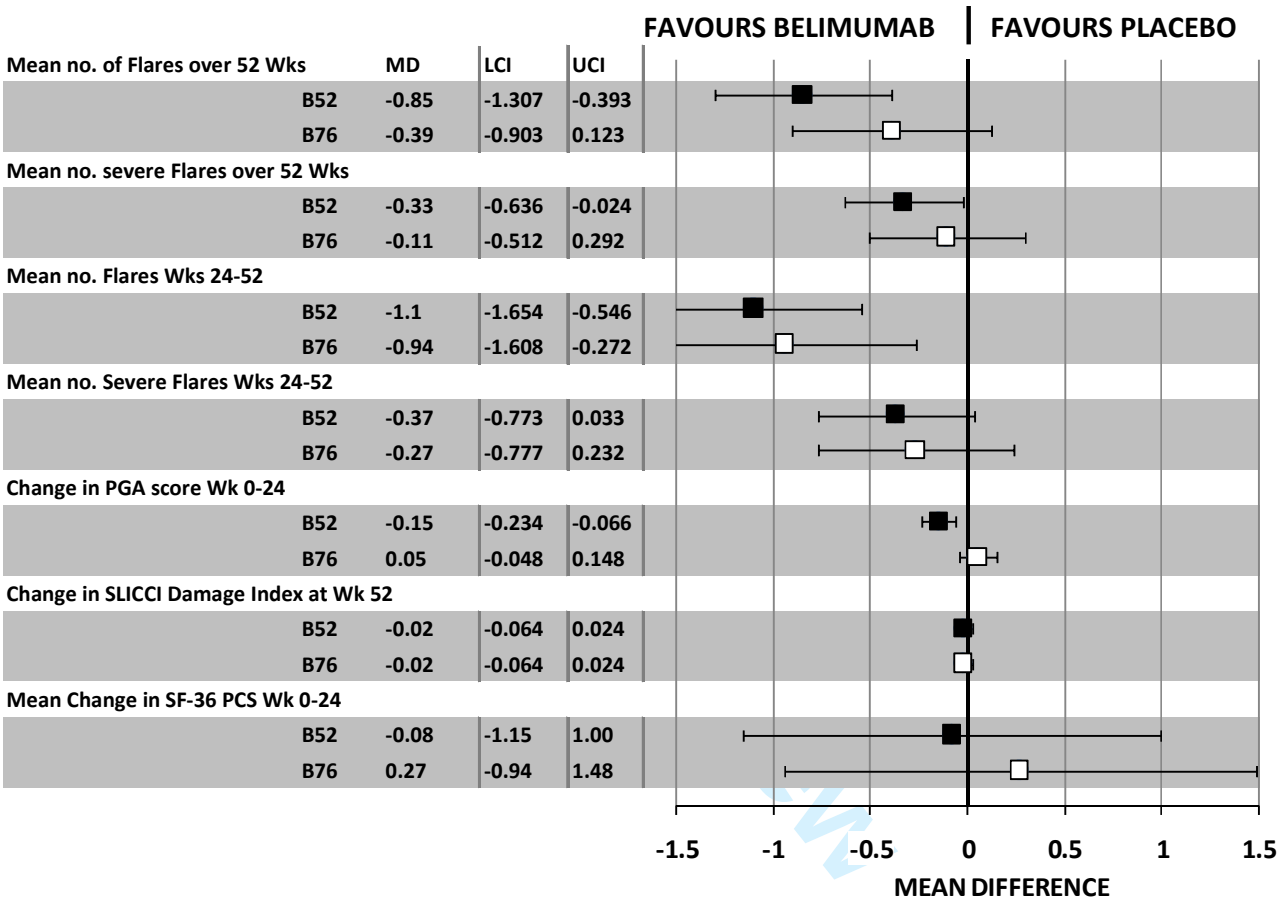
Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	Page 5-6
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	NA
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	Fig 2
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	Page 8-12
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	Page 8-12
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Figures
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	Tables
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	Discussion
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	NA
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	Discussion
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	Discussion
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	Discussion
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	End of paper

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

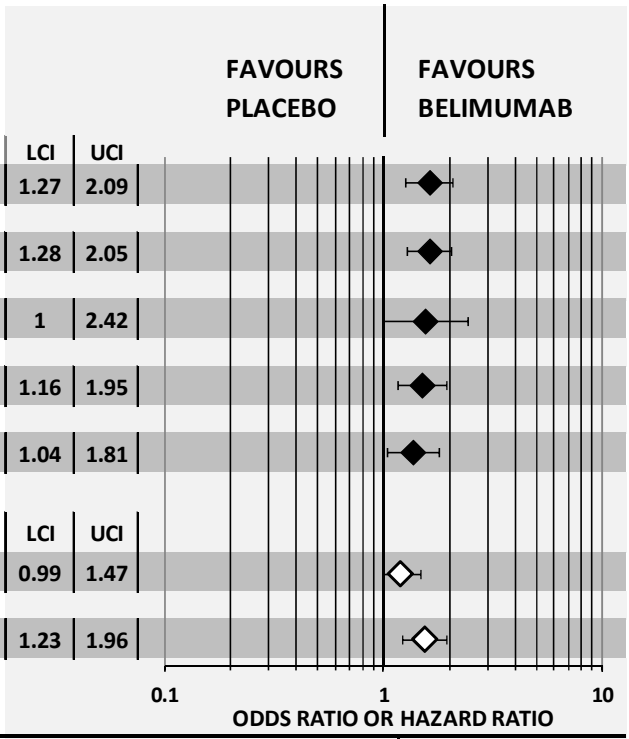
For more information, visit: **www.prisma-statement.org**.







BINARY OUTCOMES (belimumab vs. placebo)	OR	LCI	UCI
SRI wk 52	1.63	1.27	2.09
4 pnt improvement in SLEDAI score wk 52	1.63	1.28	2.05
Reduced steroid use wks 40 to 52	1.56	1	2.42
No worsening in PGA wk 52	1.5	1.16	1.95
No new 1A/2B BILAG domain scores wk 52	1.37	1.04	1.81
TIME TO EVENT (placebo vs. belimumab)	HR	LCI	UCI
TIME TO FIRST FLARE OVER 52 WKS	1.2	0.99	1.47
TIME TO FIRST SEVERE FLARE OVER 52 WEEKS	1.56	1.23	1.96



CONTINUOUS OUTCOMES	Mean diff	LCI	UCI
Mean number of Flares over 52 wks	-0.64	-1.09	-0.19
Mean number of Flares over 52 wks	-0.25	-0.45	-0.01
Mean number of Flares wks 24 - 52	-1.03	-1.46	-0.61
Mean number of severe Flares wks 24 - 52	-0.33	-0.64	-0.02
Change in PGA score week 52	-0.05	-0.25	0.14
Change in SLICC index week 52	-0.02	-0.05	0.01
Mean change SF-36 PCS	-0.1	-0.99	0.8

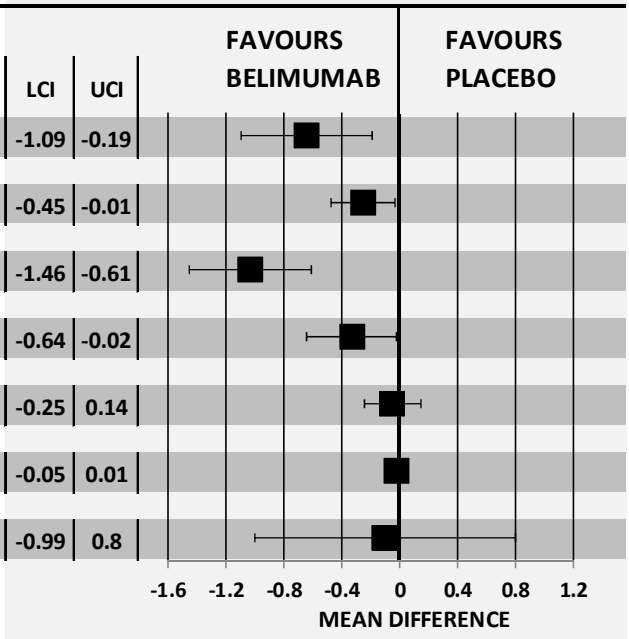


Table 1 Quality assessment of the included trials

	L02	BLISS-52	BLISS-76
Does reporting suggest that randomisation was carried out appropriately?	Yes	Yes	Yes
Does reporting suggest that the concealment of treatment allocation adequate?	Unclear	Unclear	Unclear
Were the groups reported as similar at the outset of the study in terms of prognostic factors?	Yes	Yes	Yes
Were the care providers, participants and outcome assessors reported as blind to treatment allocation?	Yes	Yes	Yes
Were any unexpected imbalances in drop-outs reported between groups?	No	No	No
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No	No	No
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	Yes	Yes

Based on the Centre for Reviews and Dissemination (2008) Systematic reviews. CRD guidance for undertaking reviews in health care. York: Centre for Reviews and Dissemination[19]

Table 2: Major characteristics of included studies

Study	Treatment (IV)	N	Mean Age (SD) yrs	SELENA-SLEDAI at entry	Geographical distribution of patients	Ethnic make-up of trial participants			Number and location of study centres
L02 2006 Phase II 52 week	Bel 1 mg/kg Bel 4 mg/kg Bel 10 mg/kg Placebo	114 111 111 113	42 (11)	> 4 points	US (98%), Canada (2%)	Caucasian	NR*	69.9%	59 North America
						African American	NR*	24.7%	
						Latino	NR*	18.5%	
BLISS-52 Phase III 52 week	Bel 1 mg/kg Bel 10 mg/kg Placebo	288 290 287	36 (11)	> 6 points	Latin America (50%), Asia (38%), Eastern Europe & Australia (13%)	Caucasian	229	27%	90 in Pacific Asia. 11 in South America & Eastern Europe
						Asian	327	38%	
						Black/African Am	30	4%	
						Alaskan Native / American Indian	279	32%	
						Native Hawaiian / Pacific Islands	0	0%	
						Multiracial	5	1%	
BLISS-76 Phase III 76 week	Bel 1 mg/kg Bel 10 mg/kg Placebo	271 273 275	40 (12)	> 6 points	US & Canada (53%), Western Europe (25%), Eastern Europe (11%), Latin America (11%)	Caucasian	569	70%	136 in North America & Europe

NR*=numbers not reported

Table 3: Outcomes reported for the BLISS-52 and 76 trials

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

*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 Erythematous Disease Activity Index; SF-36 = Short Form 36-Item Health Survey; SLICC = Systemic Lupus International Collaborating Clinics; ACR = American College of Rheumatology. Continuous outcomes are in *italics*.

Belimumab: a technological advance for SLE patients? Report of a systematic review and meta-analysis

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2013-002852.R1
Article Type:	Research
Date Submitted by the Author:	04-Jun-2013
Complete List of Authors:	Kandala, Ngianga-Bakwin; University of Warwick, Warwick Medical School Connock, Martin; University of Warwick, Division of Health Sciences, Warwick Medical School Grove, Amy; University of Warwick, Division of Health Sciences, Warwick Medical School Sutcliffe, Paul; University of Warwick, Division of Health Sciences, Warwick Medical School Mohiuddin, Syed; University of Warwick, Division of Health Sciences, Warwick Medical School Hartley, Louise; University of Warwick, Division of Health Sciences, Warwick Medical School Court, Rachel; Warwick University, Division of Health Sciences Cummis, Ewen; McMDCLtd, UK, G12 9TJ, McMaster Development Consultants Gordon, Caroline; University of Birmingham, School of Immunity and Infection, College of Medical and Dental Sciences Clarke, Aileen; University of Warwick, Division of Health Sciences
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Epidemiology, Evidence based practice, Public health
Keywords:	CLINICAL PHARMACOLOGY, EPIDEMIOLOGY, PREVENTIVE MEDICINE, PUBLIC HEALTH

Belimumab: a technological advance for Systemic Lupus Erythematosus **patients?**
Report of a systematic review and meta-analysis

Ngianga-Bakwin Kandala, Principal Research Fellow[1, 5]
Martin Connock, Senior Research Fellow[1]
Amy Grove, Project Manager[1]
Paul Sutcliffe, Associate Professor[1]
Syed Mohiuddin, Research Fellow[2]
Louise Hartley, Research Associate[1]
Rachel Court, Information Specialist[1]
Ewen Cummins, Health Economist[3]
Caroline Gordon, Professor of Rheumatology and Consultant Rheumatologist[4]
Aileen Clarke, Professor of Public Health & Health Services Research[1]

[1]Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry, UK, CV4 7AL
[2]The University of Manchester, Health Sciences, Manchester, UK, M13 9PL
[3]McMDCLtd, McMaster Development Consultants, Glasgow, UK, G12 9TJ
[4]School of Immunity and Infection, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK, B15 2TT
[5]KEMRI-University of Oxford-Wellcome Trust Collaborative Programme, Malaria Public Health and Epidemiology Group, Centre for Geographic Medicine, Nairobi, Kenya.

Correspondence to:
Ngianga-Bakwin Kandala, PhD
Principal Research Fellow in Health Technology Assessment
Warwick Evidence and Populations Evidence and Technologies
Division of Health Sciences
Warwick Medical School
University of Warwick
Coventry, CV4 7AL
Tel: +44 (0) 24761 50541
Email: N-B.Kandala@warwick.ac.uk

Short Title:

Systematic review on belimumab for SLE

INTRODUCTION

Systemic Lupus Erythematosus (SLE) is an auto-immune disease subject to relapse and remission. Incidence is estimated at between 1.0 and 10.0 per hundred thousand person years using different measures, and prevalence at between 20-70 per 100,000.[1;2] SLE is a complex multi-organ disease with a number of different manifestations.[3] Patients almost always have fatigue, often have skin rashes and arthritis and there is a wide variety of other problems which the disease can cause.

The American College of Rheumatology has defined 11 classification criteria, including: rash; photosensitivity; oral ulcers; arthritis; serositis; renal and neurological disorder.[4;5] Assessment of the patient can be difficult, as flares of the disease have to be distinguished from its complications, from comorbidity especially infection, and from adverse effects of medications.[6] SLE is more common in women (in most studies 90% or more of cases are women[2]) and in those from black and other ethnic groups. Recently age-adjusted incidence rates have been produced showing that rates are highest in women aged 40 years and over.[7] Mortality rates show that five year survival is high, at over 90%[8;9] and an overall SMR has been calculated as 2.4.[10]

Antinuclear antibodies are present in virtually all patients with SLE.[11] Anti-ds DNA antibodies are present in 50-60% patients at some point in their disease but often transiently with active disease.[11] Corticosteroids are the mainstay of treatment, they suppress disease but may themselves cause organ damage. The aim of treatment is to maintain normal function whilst suppressing disease activity and preventing organ damage,[6] achieving these conflicting aims can be difficult. Other drugs used include antimalarials such as hydroxychloroquine, and immunosuppressive drugs such as azathioprine and mycophenolatemofetil. More recently rituximab (a monoclonal antibody which reacts with the CD20 antigen expressed on B cells) has also been used, although the largest trial undertaken to date failed to reach its end point.[12]

Belimumab (Benlysta®) is an IgG1monoclonal antibody which inhibits the activity of the soluble cytokine BLyS (B lymphocyte stimulator; also known as BAFF).[13] In contrast to earlier SLE treatments, belimumab is targeted at the fundamental pathology of SLE and has been widely interpreted as representing a step change in treatment options.[13]

Belimumab was recently licensed in the USA and in Europe for treatment of autoantibody-positive SLE and is the first drug to be so licensed for several decades. The European indication is for severely affected SLE patients with active, autoantibody-positive disease and a high degree of disease activity exemplified by positive anti-ds DNA and low complement despite standard therapy.[13] Belimumab is administered by IV infusion recommended at 10 mg belimumab / kg on days 0, 14 and 28, and at 28 day intervals thereafter. A course of belimumab treatment for a 64 kg patient using the US list price of $1,477 (£926.37) for a 400 mg vial[14] would be $56,527 (£35,454) per year, and according to the US average whole sale price of $4.432 (£2780) / 400 mg vial)[15] would be $42,545 (£26,684) per year.

A number of clinical measures have been developed for tracking the progression of SLE[16] and for estimating the effects of treatment.[17] They include the Physician's Global Assessment (PGA), the SELENA-SLEDAI (Safety of Estrogen in Lupus National Assessment-Systemic Lupus Erythematosus Disease Activity Index), the BILAG Index (British Isles Lupus Assessment Group Index), and the SRI index (SLE Response Index). Their major features are summarised in Figure 1. Their complexity means that outside specialised centres they may not be widely used in routine clinical practice. The multiplicity of SLE manifestations and of the systems developed to measure them has resulted in a proliferation of outcome measures that can be reported in trials of interventions for SLE. This in turn means that by chance at least some outcome measures will generate favourable results for an intervention; hence the US Federal Drug agency (FDA) in conjunction with belimumab-trialists developed the SRI aimed at guarding against the possibility that worsening in overall disease might be masked by apparent improvement in a more narrowly defined manifestation.

[Insert Figure 1 here]

Our objective was to synthesise findings from randomised controlled trials (RCTs) of belimumab for patients with SLE and anti-nuclear and /or anti-ds DNA autoantibodies, to make an overall assessment of the performance of this drug in relation to comparator treatments using the SRI and other outcomes (as listed in Figure 1) and to assess the findings of trials in the light of population samples and geographical factors.[18]

METHODS

The study was undertaken as part of work for the National Institute for Health Research, Health Technology Assessment programme (Grant funding reference 10/73/01. Further information is available from:www.hta.ac.uk/).

Search scope

We searched for RCTs investigating belimumab administered i.v. for patients with SLE and anti-nuclear and /or anti-ds DNA autoantibodies. Comparators considered were belimumab versus placebo and belimumab versus best supportive care. Outcomes included all disease-related or health-status-related measures. There was no publication year restriction, but the search was restricted to English language references only.

Search strategy

The following eight databases were searched: Cochrane Database of Systematic Reviews; the Cochrane Central Register of Controlled Trials (CENTRAL); DARE; EMBASE; HTA Database; Medline; Pre-Medline and Science Citation Index. Search strategies for these databases used a combination of terms related to the population and interventions listed above; the specific search strategies are provided in Appendix 1. In Medline and EMBASE the subject strategies were combined with search strategies designed to identify RCTs. (Appendix 1).

Unpublished studies were identified using: Clinical Trials, Current Controlled Trials, EU Clinical Trials Register, UK Clinical Research Network Study Portfolio, National Research Register, WHO Clinical Trials Search Portal, NHS Evidence, Conference Proceedings Citation Index -Science and Google.

In addition, specific websites were searched: Medicines and Healthcare products Regulatory Agency (MHRA), European Medicines Agency (EMEA), US Food and Drug Administration (FDA) and the following specific conference proceedings: American College of Rheumatology, British Society of Rheumatology and the European League Against Rheumatism (EULAR).

Inclusion criteria: Publications were included if they described results from RCTs of belimumab for SLE patients with positive autoantibodies. Two reviewers independently assessed retrieved publications for inclusion. There were no disagreements between reviewers.

Date extraction: Potentially relevant publications were obtained in full text and assessed by the same two reviewers. One reviewer extracted data for all specified primary and secondary outcome measures, for adverse events and deaths. A second reviewer checked extracted data.

Quality evaluation: Quality assessment and risk of bias was guided by the Centre for Reviews and Dissemination (CRD) checklist[19] based on all information in the included publications which specifies reporting of randomisation, concealment of allocation, group balance, blinding, drop-outs, outcome reporting bias, and whether intention to treat analysis was used.

Statistical analysis: Unadjusted odds ratios (ORs) and mean differences were calculated for binary and continuous outcomes respectively. Statistical heterogeneity was calculated using the I^2 statistic.[20;21] There were too few studies for an analysis of publication bias.[21] Adjusted outcome measures were tabulated where these were reported. A random effects meta-analysis[22] was undertaken using the DerSimonian Laird method in STATA version 11..[23] All graphs were prepared in Microsoft Excel 2010.

RESULTS

Characteristics of included studies

We identified three placebo controlled RCTs of belimumab versus standard care: the phase III trials termed BLISS-52[24] and BLISS-76[25] and a phase II trial (study L02).[26] The PRISMA flow chart shows the process of identification of publications (see Figure 2). We identified an on-going trial in Asia.[27] All three completed trials appeared to be of good quality; however details of allocation concealment were meagre (Table 1).

[Insert Table 1 here]

[Insert Figure 2 here]

BLISS-52,[24] BLISS-76[25] and study L02[26] have been published in peer reviewed journals, however the fullest accounts in the public domain are in the FDA licensing approval documents[28;29] and the manufacturer's 2011 submission to the UK National Institute of Health and Clinical Excellence (NICE).[30] Each of these placebo-controlled randomised trials

was designed with multiple randomised groups. In the L02 trial, patients received 1 or 4 or 10 mg/kg of belimumab or placebo, while in the BLISS trials the belimumab dose regimens were 1mg/kg or 10 mg/kg. Both US and European licensing is for the 10mg/kg dose regimen. In this article we focus on efficacy results for the 10mg/kg licensed regimen relative to placebo. We also consider the off licence 1 mg/kg and 4 mg/kg dose regimens for investigation of adverse events.

Centralised, stratified randomisation was used in all three trials and arms were generally well balanced. For the phase III trials, stratification was undertaken according to race, baseline proteinuria and disease activity score (SELENA SLEDAI); for the phase I study, disease activity only was used as a stratification factor. All three trials recruited predominantly female patients (~90%) and were described as double blind. The two BLISS studies were conducted according to similar protocols.

There were differences in geographical distribution of the study centres and in the resulting ethnic racial make-up of the study populations (Table 2 and Figure 3). Thus in BLISS-76, 70% were Caucasian, 13% native American and 3% Asian, respectively, whereas in BLISS-52, 27% were Caucasian, 32% native American and 38% were Asian. Table 3 lists the major protocol pre-specified outcomes in the BLISS trials.

There were additional population differences between BLISS and L02 trials at recruitment. Reporting of results for patients with anti-nuclear and /or anti-ds DNA autoantibodies in L02 was only included for a post-hoc subgroup and primary outcomes measured in L02 were not comparable with those of the BLISS studies. For these reasons, L02 study results are included here only with regard to safety outcomes. For the BLISS trials a composite novel primary outcome measure was developed *a priori* from discussions between the FDA and the manufacturer and termed the SLE Response Index (SRI) (see Figure 1 and Table 3). The protocol pre-specified primary end point was the proportion of SRI responders at week 52. This is taken as the primary outcome in this systematic review.

[Insert Table 2 here]

[Insert Figure 3 here]

[Insert Table 3 here]

[Insert Figure 4here]

Efficacy results in the two BLISS trials for major binary effectiveness outcomes including the time to first SLE flare and to first severe flare are summarised in Figure 4. ORs have been calculated using the proportions of patients with and without events reported in the journal articles for these trials.[24;25] Safety outcomes shown in Figure 4 were calculated after pooling the number of events across the three trials (L02, BLISS-52 and BLISS-76) and are taken from the FDA documents. The hazard ratios (HRs) for time to flares were poorly reported in journal articles and the data presented are taken from the manufacturer's submission to the FDA.[28;29] As shown in Figure 4 both trials satisfied this primary end point with a better result for BLISS-52. The difference in percentage responders in the belimumab group relative to placebo group was larger in BLISS-52 (14%), than in in BLISS-76 (9.4%).

For the other binary effectiveness outcomes, the BLISS-52 trial delivered results which were more favourable to belimumab than did BLISS-76, with the latter results failing to reach a conventional level of statistical significance except for the ≥ 4 point improvement in SLEDAI score at week 52. The journal articles and manufacturer's submissions to the FDA and to NICE used a logistic regression model and reported ORs adjusted according to the stratification factors employed at randomisation. Adjusted ORs 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 = 0.0207). Again a superior response was found for the BLISS-52 trial. By week 76, the unadjusted OR for the SRI response in the BLISS-76 trial ceased to reach statistical significance (Figure 4); this also held for the reported OR adjusted by logistic regression (OR 1.31, 95% CI: 0.92 – 1.87, p = 0.1323).[29]

With regard to time to first flare or to first severe flare (each estimated over 52 weeks follow up) the responses reported in the FDA submission are again superior for BLISS-52. Each outcome failed to reach conventional statistical significance for BLISS-76. The FDA submission additionally reported more mature results estimated over 76 weeks of follow up for BLISS-76, and again these indicate lack of statistical significance for both outcomes (HR for first flare: 1.05, 95% CI: 0.88 – 1.27; HR for first severe flare 1.30, 95% CI: 0.92 – 1.85).

Figure 4 shows the results for major safety outcomes. Although there were more serious adverse events, more serious infections and more deaths associated with belimumab than with placebo, none of the ORs for these outcomes reached statistical significance. There were 14 deaths during the controlled phase of the three trials; three in the placebo group (n=675), and 11 in the belimumab groups (n=1458) with six in the 10mg/kg and five in the

1mg/kg groups, respectively (odds ratio 11.7; 95% CI 0.474 to 6.124). The causes of death were various: five were infection-related, three were strokes, three cardiovascular events, two suicides, one cancer, one from SLE-related complications, and two were of unknown cause.

Results for continuous outcomes are summarised in Figure 5. Mean changes from baseline reported in the BLISS journal articles and in the manufacturer's submissions to the FDA and NICE have been used to generate a mean difference statistic (sometimes termed "weighted mean difference"[31]). These revealed superiority of response in BLISS-52 relative to BLISS-76 for all reported outcomes, a pattern similar to that for binary outcomes. Mean changes from baseline for FACIT-fatigue scores and for EQ-5D utility scores (not pictured) did not reach statistical significance and again improvement seen in BLISS-52 for these was superior to that seen in BLISS-76.

In summary, BLISS-52 showed a systematic superiority over BLISS-76 in apparent benefit of belimumab across the full range of effectiveness outcomes (binary, time to event and continuous), which may reflect geographical differences between the trials (Table 2 and Figure 3). The primary outcome in BLISS-76 was achieved (adjusted OR 1.52, 95% CI 1.07 to 2.15) but large geographical differences within BLISS-76 were striking: rates of 32% (46 out of 145), and 35% (47 out of 136), for placebo and belimumab respectively, were reported for patients from North America and Canada (a < 3% greater response for belimumab), whereas for BLISS-76 patients outside these regions a > 15% greater response for belimumab over placebo was reported, 71 of 137 (51.8%) for belimumab and 47 of 130 (36.1%) for placebo. In comparison, the corresponding rates for patients from Latin America in BLISS-52 were 49% placebo (71 out of 145), and 61% belimumab (85 out of 140).

[Insert Figure 5 here]

The manufacturer's submissions to the FDA and to NICE combined results from the two BLISS trials by pooling the patients and applying the logistic regression model described above; for the primary outcome (proportion of SRI responders at week 52), the difference between the belimumab and placebo groups was 11.8%.[28]

An alternate method of combining trials by meta-analysis of study level results from the two BLISS trials showed a statistically significant benefit of belimumab for most main outcomes including SRI, SELENA-SLEDAI, worsening in PGA, steroid use, BILAG score or, time to first severe flare, and mean number of flares and severe flares over 52 weeks and weeks 24

to 52 (Figure 6). Tests for statistical heterogeneity of ORs and HR outcomes were not significant. These results, and those from pooling individual patient data from the two trials prior to logistic regression, mask the systematic difference between trials across all outcomes.

[Insert Figure 6 here]

DISCUSSION

We undertook a systematic review of the clinical effectiveness of belimumab, a new treatment targeted at systemic lupus erythematosus (SLE) patients with anti-nuclear and /or anti-ds DNA autoantibodies. We performed an extensive search and systematic review of both completed and on-going trials using a number of databases and by checking reference lists. Data were extracted independently and studies were quality assessed. Random effects meta-analysis was undertaken.

We identified three RCTs (L02, BLISS-52, BLISS-76) reporting data on over 2000 patients. In contrast to the BLISS trials, L02 recruited patients who were not necessarily current carriers of anti-nuclear or anti ds DNA antibodies at study commencement. L02 failed to demonstrate clinical effectiveness for its primary end points.[26] Meta-analysis of the BLISS studies showed a benefit of belimumab with the main primary outcome (SRI), showing improvement at 52 weeks (OR 1.63; 95% CI: 1.27-2.09 p<0.001) although by week 76, the proportion of SRI responders in the BLISS-76 trial ceased to reach statistical significance (OR 1.31; 95% CI: 0.92–1.87 p=0.1323). There were no significant differences between placebo and intervention groups in quality of life or adverse events.

We found that the benefits of belimumab were systematically greater across the board (although not significantly so) in the BLISS-52 trial and although tests for statistical heterogeneity were negative, geographical location of study centres and the racial background and ethnicity of participants varied considerably. If the two BLISS trials were drawn from the same underlying populations, whilst one might expect outcomes to differ, we would anticipate that this would occur randomly between trials– some better some worse than the other.

A few studies have directly assessed the existence of and importance of geographical differences in trial outcomes.[32-34] Key factors contributing to such differences are variation in underlying patient population characteristics and variation in study execution. Vickers et al,[33]

found that Eastern Asian and Eastern European studies had a higher proportion of positive trial results when compared to other countries. This is seen in the present case for the primary outcome where both the belimumab and placebo response rates in BLISS 52 study were greater than those in BLISS-76 and, remarkably, the placebo response rate in BLISS-52 (49%) was greater than that for the belimumab arm of BLISS-76 (43%). O'Shea and DeMets also report that within the Beta-Blocker Heart Attack Trial (BHAT), not only was there a difference in the direction, but also in the size of treatment effect between Canada and the US, although it should be noted that the original aim of that trial was not investigation of international differences in treatment effect.[35] One study found that 96-99% of the total variance in the "*Global utilisation of strategies to open occluded coronary arteries IV acute coronary syndromes*" (GUSTO IV ACS) trial could be accounted for by patient-level factors.[36]

International trials need to harmonise training of investigators, patient selection, treatment management, thresholds to centre admission, access to facilities, ascertainment of endpoints and, by implication, results of interest [37-44] since it is possible that in centres in different countries these factors may differ systematically.[37] Equally, underlying differences in populations and countries (ethnicity, genetics, socio-economic status and health-care systems), and the nature and epidemiology of SLE according to ethnic background may result in differences in reporting of outcomes and pooled results.

The outcomes used in the BLISS trials would be unfamiliar to most of the investigators and it is possible that criteria may have differed between countries. In particular the Physician Global Assessment (PGA) is an important element of the outcomes measured (see Figure 1). PGA was measured as an outcome in itself, and it is also incorporated in SRI. PGA is of concern because as a global physician assessment of a patient's SLE status, it is subjective. The investigators reported a nearly 10% difference between the BLISS-52 and BLISS-76 studies in estimates of percentage change in PGA score in intervention groups at week 24 compared to baseline and this single result in one of the two trials is likely to have had an important influence on findings of the effectiveness of belimumab in SLE patients.

The latest results of belimumab in patients with SLE (phase II study design, uncontrolled extension study) reported that of 449 patients with active SLE (USA/Canada) 177 (39.4%) patients remained on treatment after 7 years of therapy (i.e. approximately 1746 cumulative patients-years) and that this subgroup exhibited durable sustained improvement in SLE disease activity (SRI and PGA).[30]

CONCLUSIONS

In conclusion, systematic review and random effects meta-analysis of two RCTs of belimumab for patients with autoantibody positive SLE demonstrated positive results in the main outcome at week 52. However, in view of the different populations studied at different locations in the BLISS trials and the consistently superior results from one trial compared to the other, we consider that population heterogeneity, geographical differences and variation in trial conduct and outcome assessment, may have played a role in influencing outcomes. However the generalisability of results pooled meta-analytically or by logistic regression should be viewed with caution and we suggest that it is too early to draw strong conclusions in this case.

ARTICLE FOCUS

- SLE is a complex multi-organ auto-immune disease subject to relapse and remission.
- Patients almost always have fatigue, skin rashes and arthritis and there is a wide variety of other problems which the disease can cause.
- Belimumab is a new treatment specifically targeted against SLE.

KEY MESSAGES

1. Combining the results from two RCTs suggests that belimumab is clinically effective for SLE patients.
2. However, all outcomes were systematically superior in one trial compared with the other.
3. Different trial conduct and populations mean that it is too early to draw generalisable conclusions.

STRENGTHS AND LIMITATIONS

- At first sight combined meta analytic evidence suggests that belimumab is clinically effective for patients with severe SLE.
- We suggest that it is too early to draw strong conclusions because the two relevant trials cover different populations in different countries and there may be differences in trial conduct and outcome assessment.

Acknowledgements

The authors would like to thank the National Institute for Health Research, Health Technology Assessment programme for funding this work.

Funding statement

This work was supported by the National Institute for Health Research, Health Technology Assessment programme [grant number 10/73/01].

Competing interest statement

No conflicts of interest.

Contributions:

N-BK: Conception and design. Data analysis and interpretation. Drafting the article. Critical revisions for important intellectual content. Approval of final article for submission.

MC: Conception and design. Data analysis and interpretation. Literature review. Interpretation of results. Drafting the article. Critical revisions for important intellectual content. Approval of final article for submission.

AG: Interpretation of results. Critical revisions for important intellectual content.

PS: Literature review. Interpretation of results. Critical revisions for important intellectual content.

SM: Data analysis and interpretation. Interpretation of results. Critical revisions for important intellectual content.

LH: Literature review. Interpretation of results. Critical revisions for important intellectual content.

RC: Literature review. Critical revisions for important intellectual content.

EC: Interpretation of results. Critical revisions for important intellectual content.

CG: Interpretation of results. Critical revisions for important intellectual content.

AC: Conception and design. Interpretation of results. Drafting the article. Critical revisions for important intellectual content. Approval of final article for submission.

All authors read and approved the final manuscript.

Reference List

(1) Bernatsky S, Joseph L, Pineau CA, Tamblyn R, Feldman DE, Clarke AE. A population-based assessment of systemic lupus erythematosus incidence and prevalence--results and implications of using administrative data for epidemiological studies. *Rheumatology (Oxford)* 2007; 46(12):1814-1818.

(2) Pons-Estel GJ, Alarcon GS, Scofield L, Reinlib L, Cooper GS. Understanding the epidemiology and progression of systemic lupus erythematosus. *Semin Arthritis Rheum* 2010; 39(4):257-268.

(3) Gordon C, Li CK, Isenberg DA. Systemic lupus erythematosus. *Medicine* 2010; 38(2):73-80.

(4) Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1997; 40(9):1725.

(5) Tan EM, Cohen AS, Fries JF, Masi AT, McShane DJ, Rothfield NF et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1982; 25(11):1271-1277.

(6) Griffiths B, Mosca M, Gordon C. Assessment of patients with systemic lupus erythematosus and the use of lupus disease activity indices. *Best Practice and Research in Clinical Rheumatology* 2005; 19(5 SPEC. ISS.):685-708.

(7) Somers EC, Thomas SL, Smeeth L, Schoonen WM, Hall AJ. Incidence of systemic lupus erythematosus in the United Kingdom, 1990-1999. *Arthritis Rheum* 2007; 57(4):612-618.

(8) Pons-Estel GJ, Alarcon GS, McGwin G, Jr., Danila MI, Zhang J, Bastian HM et al. Protective effect of hydroxychloroquine on renal damage in patients with lupus nephritis: LXV, data from a multiethnic US cohort. *Arthritis Rheum* 2009; 61(6):830-839.

(9) Pons-Estel GJ, Gonzalez LA, Zhang J, Burgos PI, Reveille JD, Vila LM et al. Predictors of cardiovascular damage in patients with systemic lupus erythematosus: data from LUMINA (LXVIII), a multiethnic US cohort. *Rheumatology (Oxford)* 2009; 48(7):817-822.

(10) Bernatsky S, Boivin JF, Joseph L, Manzi S, Ginzler E, Gladman DD et al. Mortality in systemic lupus erythematosus. *Arthritis and Rheumatism* 2006; 54(8):2550-2557.

(11) Kurien BT, Scofield RH. Autoantibody determination in the diagnosis of systemic lupus erythematosus. *Scand J Immunol* 2006; 64(3):227-235.

(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; 62(1):222-233.

(13) Summary of product characteristics. 2012 Available from: URL: ema.europa.eu/Find medicine/Human medicines/European Public Assessment Reports Accessed January 2012

(14) American College of Rheumatology. Belimumab for Systemic Lupus Erythematosus. 2011 Available from: URL: http://www.rheumatology.org/publications/hotline/2011_03_15_belimumab.asp Accessed January 2012.

(15) New Drug Bulletin. Belimumab (Benlysta - Human Genome Sciences, Inc). 2012 Available from: URL:http://healthcare.utah.edu/pharmacy/bulletins/NDB_228.pdf Accessed January 2012.

(16) Isenberg DA, Rahman A, Allen E, Farewell V, Akil M, Bruce IN et al. BILAG 2004. Development and initial validation of an updated version of the British Isles Lupus Assessment Group's disease activity index for patients with systemic lupus erythematosus. *Rheumatology* 2005; 44(7):902-906.

(17) Gordon C, Bertsias G, Ioannidis JPA, Boletis J, Bombardieri S, Cervera R et al. EULAR points to consider for conducting clinical trials in Systemic lupus erythematosus. *Ann Rheum Dis* 2009; 68(4):470-476.

(18) Burgos PI, McGwin G, Jr., Pons-Estel GJ, Reveille JD, Alarcon GS, Vila LM. US patients of Hispanic and African ancestry develop lupus nephritis early in the disease course: data from LUMINA, a multiethnic US cohort (LUMINA LXXIV). *Ann Rheum Dis* 2011; 70(2):393-394.

(19) Khan KS, Ter Riet G, Glanville J, Snowdon AJ, Kleijnen J. Reviews of Research on Effectiveness. CRD's Guidance for Carrying Out or Commissioning Reviews. 2 nd Edition. (ISBN 1900640201) CRD Report No. 4. York: NHS Centre for Reviews and Dissemination (CRD), University of York. 2001 Available from: URL:http://www.york.ac.uk/inst/crd/report4.htm; Accessed January 2012.

(20) Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327(7414):557-560.

(21) Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. 2012 Available from: URL:www.cochrane-handbook.org Accessed January 2012.

(22) DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; 7(3):177-188.

(23) STATA Statistical analysis and software. 2011 Available from: URL:http://www.stata.com/ Accessed April 2012.

(24) 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(9767):721-731.

(25) Furie R, Petri M, Zamani O, Cervera R, Wallace DJ, Tegzova D et al. A phase III, randomized, placebo-controlled study of belimumab, a monoclonal antibody that inhibits B lymphocyte stimulator, in patients with systemic lupus erythematosus. *Arthritis Rheum* 2011; 63(12):3918-3930.

(26) 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(9):1168-1178.

(27) GlaxoSmithKline. GSK1550188 A 52 week study of belimumab versus placebo in the treatment of subjects with SLE located in North East Asia. 2011 Available from: URL:http://clinicaltrials.gov/ct2/show/NCT01345253 Accessed January 2012.

(28) 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. 2010 Available from: URL:http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/ArthritisDrugsAdvisoryCommittee/UCM233581.pdf Accessed January 2012.

(29) U.S.Food and Drugs Administration. FDA briefing package. 2010 Available from: URL:http://www.fda.gov/downloads/advisorycommittees/committeesmeetingmaterials/drugs/arthritisdrugsadvisorycommittee/ucm233579.pdf Accessed January 2012.

(30) GlaxoSmithKline. Single Technology Appraisal (STA): Belimumab for the treatment of active autoantibody-positive systemic lupus erythematosus. 2011 Available from: URL:http://www.guidance.nice.org.uk/TA/live/13307/56490/56490.pdf Accessed January 2012.

(31) Egger M, Smith GD, Altman DG. Systematic Reviews in Health Care, Meta-analysis in context. 2 ed. BMJ Books; 2001.

(32) Pan Z, Trikalinos TA, Kavvoura FK, Lau J, Ioannidis JP. Local literature bias in genetic epidemiology: an empirical evaluation of the Chinese literature. *PLoS Med* 2005; 2(12):e334.

(33) Vickers A, Goyal N, Harland R, Rees R. Do certain countries produce only positive results? A systematic review of controlled trials. *Control Clin Trials* 1998; 19(2):159-166.

(34) Zhang D, Freemantle N, Cheng KK. Are randomized trials conducted in China or India biased? A comparative empirical analysis. *J Clin Epidemiol* 2011; 64(1):90-95.

(35) O'Shea JC, DeMets DL. Statistical issues relating to international differences in clinical trials. *Am Heart J* 2001; 142(1):21-28.

(36) Chang WC, Midodzi WK, Westerhout CM, Boersma E, Cooper J, Barnathan ES et al. Are international differences in the outcomes of acute coronary syndromes apparent or real? A multilevel analysis. *J Epidemiol Community Health* 2005; 59(5):427-433.

(37) White HD. International differences: selection, noise, or real? *Eur Heart J* 2000; 21(5):339-342.

(38) Fox KA, Goodman S, Bigonzi F, Le L, V, Cohen M. Inter-regional differences and outcome in unstable angina; analysis of the international ESSENCE trial. Efficacy and Safety of Subcutaneous Enoxaparin in Non-Q-wave Coronary Events. *Eur Heart J* 2000; 21(17):1433-1439.

(39) Sinnaeve PR, Van de Werf FJ. Global patterns of health care for acute coronary syndromes. *Curr Opin Cardiol* 2004; 19(6):625-630.

(40) Faergeman O, Kjekshus J, Cook T, Pyorala K, Wilhelmsen L, Thorgeirsson G et al. Differences in the treatment of coronary heart disease between countries as revealed in the Scandinavian Simvastatin Survival Study (4S). *Eur Heart J* 1998; 19(10):1531-1537.

(41) Berger JS, Stebbins A, Granger CB, Ohman EM, Armstrong PW, Van de Werf F et al. Initial aspirin dose and outcome among ST-elevation myocardial infarction patients treated with fibrinolytic therapy. *Circulation* 2008; 117(2):192-199.

(42) Welsh RC, Chang W, Goldstein P, Adgey J, Granger CB, Verheugt FW et al. Time to treatment and the impact of a physician on prehospital management of acute ST elevation myocardial infarction: insights from the ASSENT-3 PLUS trial. *Heart* 2005; 91(11):1400-1406.

(43) Yusuf S, Flather M, Pogue J, Hunt D, Varigos J, Piegas L et al. Variations between countries in invasive cardiac procedures and outcomes in patients with suspected unstable angina or myocardial infarction without initial ST elevation. OASIS (Organisation to Assess Strategies for Ischaemic Syndromes) Registry Investigators. *Lancet* 1998; 352(9127):507-514.

(44) Bleyer AJ, Hylander B, Sudo H, Nomoto Y, de la Torre E, Chen RA et al. An international study of patient compliance with hemodialysis. *JAMA* 1999; 281(13):1211-1213.

Belimumab: a technological advance for Systemic Lupus Erythematosus patients?

Report of a systematic review and meta-analysis

Ngianga-Bakwin Kandala, Principal Research Fellow[1, 5]
Martin Connock, Senior Research Fellow[1]
Amy Grove, Project Manager[1]
Paul Sutcliffe, Associate Professor[1]
Syed Mohiuddin, Research Fellow[2]
Louise Hartley, Research Associate[1]
Rachel Court, Information Specialist[1]
Ewen Cummins, Health Economist[3]
Caroline Gordon, Professor of Rheumatology and Consultant Rheumatologist[4]
Aileen Clarke, Professor of Public Health & Health Services Research[1]

[1]Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry, UK, CV4 7AL
[2]The University of Manchester, Health Sciences, Manchester, UK, M13 9PL
[3]McMDCLtd, McMaster Development Consultants, Glasgow, UK, G12 9TJ
[4]School of Immunity and Infection, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK, B15 2TT
[5]KEMRI-University of Oxford-Wellcome Trust Collaborative Programme, Malaria Public Health and Epidemiology Group, Centre for Geographic Medicine, Nairobi, Kenya.

Correspondence to:
Ngianga-Bakwin Kandala, PhD
Principal Research Fellow in Health Technology Assessment
Warwick Evidence and Populations Evidence and Technologies
Division of Health Sciences
Warwick Medical School
University of Warwick
Coventry, CV4 7AL
Tel: +44 (0) 24761 50541
Email: N-B.Kandala@warwick.ac.uk

Short Title:

Systematic review on belimumab for SLE

INTRODUCTION

Systemic Lupus Erythematosus (SLE) is an auto-immune disease subject to relapse and remission. Incidence is estimated at between 1.0 and 10.0 per hundred thousand person years using different measures, and prevalence at between 20-70 per 100,000.[1;2] SLE is a complex multi-organ disease with a number of different manifestations.[3] Patients almost always have fatigue, often have skin rashes and arthritis and there is a wide variety of other problems which the disease can cause.

The American College of Rheumatology has defined 11 classification criteria, including: rash; photosensitivity; oral ulcers; arthritis; serositis; renal and neurological disorder.[4;5] Assessment of the patient can be difficult, as flares of the disease have to be distinguished from its complications, from comorbidity especially infection, and from adverse effects of medications.[6] SLE is more common in women (in most studies 90% or more of cases are women[2]) and in those from black and other ethnic groups. Recently age-adjusted incidence rates have been produced showing that rates are highest in women aged 40 years and over.[7] Mortality rates show that five year survival is high, at over 90%[8;9] and an overall SMR has been calculated as 2.4.[10]

Antinuclear antibodies are present in virtually all patients with SLE.[11] Anti-ds DNA antibodies are present in 50-60% patients at some point in their disease but often transiently with active disease.[11] Corticosteroids are the mainstay of treatment, they suppress disease but may themselves cause organ damage. The aim of treatment is to maintain normal function whilst suppressing disease activity and preventing organ damage,[6] achieving these conflicting aims can be difficult. Other drugs used include antimalarials such as hydroxychloroquine, and immunosuppressive drugs such as azathioprine and mycophenolatemofetil. More recently rituximab (a monoclonal antibody which reacts with the CD20 antigen expressed on B cells) has also been used, although the largest trial undertaken to date failed to reach its end point.[12]

Belimumab (Benlysta®) is an IgG1monoclonal antibody which inhibits the activity of the soluble cytokine BLyS (B lymphocyte stimulator; also known as BAFF).[13] In contrast to earlier SLE treatments, belimumab is targeted at the fundamental pathology of SLE and has been widely interpreted as representing a step change in treatment options.[13]

Belimumab was recently licensed in the USA and in Europe for treatment of autoantibody-positive SLE and is the first drug to be so licensed for several decades. The European indication is for severely affected SLE patients with active, autoantibody-positive disease and a high degree of disease activity exemplified by positive anti-ds DNA and low complement despite standard therapy.[13] Belimumab is administered by IV infusion recommended at 10 mg belimumab / kg on days 0, 14 and 28, and at 28 day intervals thereafter. A course of belimumab treatment for a 64 kg patient using the US list price of $1,477 (£926.37) for a 400 mg vial[14] would be $56,527 (£35,454) per year, and according to the US average whole sale price of $4.432 (£2780) / 400 mg vial)[15] would be $42,545 (£26,684) per year.

A number of clinical measures have been developed for tracking the progression of SLE[16] and for estimating the effects of treatment.[17] They include the Physician's Global Assessment (PGA), the SELENA-SLEDAI (Safety of Estrogen in Lupus National Assessment-Systemic Lupus Erythematosus Disease Activity Index), the BILAG Index (British Isles Lupus Assessment Group Index), and the SRI index (SLE Response Index). Their major features are summarised in Figure 1. Their complexity means that outside specialised centres they may not be widely used in routine clinical practice. The multiplicity of SLE manifestations and of the systems developed to measure them has resulted in a proliferation of outcome measures that can be reported in trials of interventions for SLE. This in turn means that by chance at least some outcome measures will generate favourable results for an intervention; hence the US Federal Drug agency (FDA) in conjunction with belimumab-trialists developed the SRI aimed at guarding against the possibility that worsening in overall disease might be masked by apparent improvement in a more narrowly defined manifestation.

[Insert Figure 1 here]

Our objective was to synthesise findings from randomised controlled trials (RCTs) of belimumab for patients with SLE and anti-nuclear and /or anti-ds DNA autoantibodies, to make an overall assessment of the performance of this drug in relation to comparator treatments using the SRI and other outcomes (as listed in Figure 1) and to assess the findings of trials in the light of population samples and geographical factors.[18]

METHODS

The study was undertaken as part of work for the National Institute for Health Research, Health Technology Assessment programme (Grant funding reference 10/73/01. Further information is available from:www.hta.ac.uk/).

Search scope

We searched for RCTs investigating belimumab administered i.v. for patients with SLE and anti-nuclear and /or anti-ds DNA autoantibodies. Comparators considered were belimumab versus placebo and belimumab versus best supportive care. Outcomes included all disease-related or health-status-related measures. There was no publication year restriction, but the search was restricted to English language references only.

Search strategy

The following eight databases were searched: Cochrane Database of Systematic Reviews; the Cochrane Central Register of Controlled Trials (CENTRAL); DARE; EMBASE; HTA Database; Medline; Pre-Medline and Science Citation Index. Search strategies for these databases used a combination of terms related to the population and interventions listed above; the specific search strategies are provided in Appendix 1. In Medline and EMBASE the subject strategies were combined with search strategies designed to identify RCTs. (Appendix 1).

Unpublished studies were identified using: Clinical Trials, Current Controlled Trials, EU Clinical Trials Register, UK Clinical Research Network Study Portfolio, National Research Register, WHO Clinical Trials Search Portal, NHS Evidence, Conference Proceedings Citation Index -Science and Google.

In addition, specific websites were searched: Medicines and Healthcare products Regulatory Agency (MHRA), European Medicines Agency (EMEA), US Food and Drug Administration (FDA) and the following specific conference proceedings: American College of Rheumatology, British Society of Rheumatology and the European League Against Rheumatism (EULAR).

Inclusion criteria: Publications were included if they described results from RCTs of belimumab for SLE patients with positive autoantibodies. Two reviewers independently assessed retrieved publications for inclusion. There were no disagreements between reviewers.

Date extraction: Potentially relevant publications were obtained in full text and assessed by the same two reviewers. One reviewer extracted data for all specified primary and secondary outcome measures, for adverse events and deaths. A second reviewer checked extracted data.

Quality evaluation: Quality assessment and risk of bias was guided by the Centre for Reviews and Dissemination (CRD) checklist[19] based on all information in the included publications which specifies reporting of randomisation, concealment of allocation, group balance, blinding, drop-outs, outcome reporting bias, and whether intention to treat analysis was used.

Statistical analysis: Unadjusted odds ratios (ORs) and mean differences were calculated for binary and continuous outcomes respectively. Statistical heterogeneity was calculated using the I^2 statistic.[20;21] There were too few studies for an analysis of publication bias.[21] Adjusted outcome measures were tabulated where these were reported. A random effects meta-analysis[22] was undertaken using the DerSimonian Laird method in STATA version 11..[23] All graphs were prepared in Microsoft Excel 2010.

RESULTS

Characteristics of included studies

We identified three placebo controlled RCTs of belimumab versus standard care: the phase III trials termed BLISS-52[24] and BLISS-76[25] and a phase II trial (study L02).[26] The PRISMA flow chart shows the process of identification of publications (see Figure 2). We identified an on-going trial in Asia.[27] All three completed trials appeared to be of good quality; however details of allocation concealment were meagre (Table 1).

[Insert Table 1 here]

[Insert Figure 2 here]

BLISS-52,[24] BLISS-76[25] and study L02[26] have been published in peer reviewed journals, however the fullest accounts in the public domain are in the FDA licensing approval documents[28;29] and the manufacturer's 2011 submission to the UK National Institute of Health and Clinical Excellence (NICE).[30] Each of these placebo-controlled randomised trials

was designed with multiple randomised groups. In the L02 trial, patients received 1 or 4 or 10 mg/kg of belimumab or placebo, while in the BLISS trials the belimumab dose regimens were 1mg/kg or 10 mg/kg. Both US and European licensing is for the 10mg/kg dose regimen. In this article we focus on efficacy results for the 10mg/kg licensed regimen relative to placebo. We also consider the off licence 1 mg/kg and 4 mg/kg dose regimens for investigation of adverse events.

Centralised, stratified randomisation was used in all three trials and arms were generally well balanced. For the phase III trials, stratification was undertaken according to race, baseline proteinuria and disease activity score (SELENA SLEDAI); for the phase I study, disease activity only was used as a stratification factor. All three trials recruited predominantly female patients (~90%) and were described as double blind. The two BLISS studies were conducted according to similar protocols.

There were differences in geographical distribution of the study centres and in the resulting ethnic racial make-up of the study populations (Table 2 and Figure 3). Thus in BLISS-76, 70% were Caucasian, 13% native American and 3% Asian, respectively, whereas in BLISS-52, 27% were Caucasian, 32% native American and 38% were Asian. Table 3 lists the major protocol pre-specified outcomes in the BLISS trials.

There were additional population differences between BLISS and L02 trials at recruitment. Reporting of results for patients with anti-nuclear and /or anti-ds DNA autoantibodies in L02 was only included for a post-hoc subgroup and primary outcomes measured in L02 were not comparable with those of the BLISS studies. For these reasons, L02 study results are included here only with regard to safety outcomes. For the BLISS trials a composite novel primary outcome measure was developed *a priori* from discussions between the FDA and the manufacturer and termed the SLE Response Index (SRI) (see Figure 1 and Table 3). The protocol pre-specified primary end point was the proportion of SRI responders at week 52. This is taken as the primary outcome in this systematic review.

[Insert Table 2 here]

[Insert Figure 3 here]

[Insert Table 3 here]

[Insert Figure 4here]

Efficacy results in the two BLISS trials for major binary effectiveness outcomes including the time to first SLE flare and to first severe flare are summarised in Figure 4. ORs have been calculated using the proportions of patients with and without events reported in the journal articles for these trials.[24;25] Safety outcomes shown in Figure 4 were calculated after pooling the number of events across the three trials (L02, BLISS-52 and BLISS-76) and are taken from the FDA documents. The hazard ratios (HRs) for time to flares were poorly reported in journal articles and the data presented are taken from the manufacturer's submission to the FDA.[28;29] As shown in Figure 4 both trials satisfied this primary end point with a better result for BLISS-52. The difference in percentage responders in the belimumab group relative to placebo group was larger in BLISS-52 (14%), than in in BLISS-76 (9.4%).

For the other binary effectiveness outcomes, the BLISS-52 trial delivered results which were more favourable to belimumab than did BLISS-76, with the latter results failing to reach a conventional level of statistical significance except for the ≥ 4 point improvement in SLEDAI score at week 52. The journal articles and manufacturer's submissions to the FDA and to NICE used a logistic regression model and reported ORs adjusted according to the stratification factors employed at randomisation. Adjusted ORs 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 = 0.0207). Again a superior response was found for the BLISS-52 trial. By week 76, the unadjusted OR for the SRI response in the BLISS-76 trial ceased to reach statistical significance (Figure 4); this also held for the reported OR adjusted by logistic regression (OR 1.31, 95% CI: 0.92 – 1.87, p = 0.1323).[29]

With regard to time to first flare or to first severe flare (each estimated over 52 weeks follow up) the responses reported in the FDA submission are again superior for BLISS-52. Each outcome failed to reach conventional statistical significance for BLISS-76. The FDA submission additionally reported more mature results estimated over 76 weeks of follow up for BLISS-76, and again these indicate lack of statistical significance for both outcomes (HR for first flare: 1.05, 95% CI: 0.88 – 1.27; HR for first severe flare 1.30, 95% CI: 0.92 – 1.85).

Figure 4 shows the results for major safety outcomes. Although there were more serious adverse events, more serious infections and more deaths associated with belimumab than with placebo, none of the ORs for these outcomes reached statistical significance. There were 14 deaths during the controlled phase of the three trials; three in the placebo group (n=675), and 11 in the belimumab groups (n=1458) with six in the 10mg/kg and five in the

1mg/kg groups, respectively (odds ratio 11.7; 95% CI 0.474 to 6.124). The causes of death were various: five were infection-related, three were strokes, three cardiovascular events, two suicides, one cancer, one from SLE-related complications, and two were of unknown cause.

Results for continuous outcomes are summarised in Figure 5. Mean changes from baseline reported in the BLISS journal articles and in the manufacturer's submissions to the FDA and NICE have been used to generate a mean difference statistic (sometimes termed "weighted mean difference"[31]). These revealed superiority of response in BLISS-52 relative to BLISS-76 for all reported outcomes, a pattern similar to that for binary outcomes. Mean changes from baseline for FACIT-fatigue scores and for EQ-5D utility scores (not pictured) did not reach statistical significance and again improvement seen in BLISS-52 for these was superior to that seen in BLISS-76.

In summary, BLISS-52 showed a systematic superiority over BLISS-76 in apparent benefit of belimumab across the full range of effectiveness outcomes (binary, time to event and continuous), which may reflect geographical differences between the trials (Table 2 and Figure 3). The primary outcome in BLISS-76 was achieved (adjusted OR 1.52, 95% CI 1.07 to 2.15) but large geographical differences within BLISS-76 were striking: rates of 32% (46 out of 145), and 35% (47 out of 136), for placebo and belimumab respectively, were reported for patients from North America and Canada (a < 3% greater response for belimumab), whereas for BLISS-76 patients outside these regions a > 15% greater response for belimumab over placebo was reported, 71 of 137 (51.8%) for belimumab and 47 of 130 (36.1%) for placebo. In comparison, the corresponding rates for patients from Latin America in BLISS-52 were 49% placebo (71 out of 145), and 61% belimumab (85 out of 140).

[Insert Figure 5 here]

The manufacturer's submissions to the FDA and to NICE combined results from the two BLISS trials by pooling the patients and applying the logistic regression model described above; for the primary outcome (proportion of SRI responders at week 52), the difference between the belimumab and placebo groups was 11.8%.[28]

An alternate method of combining trials by meta-analysis of study level results from the two BLISS trials showed a statistically significant benefit of belimumab for most main outcomes including SRI, SELENA-SLEDAI, worsening in PGA, steroid use, BILAG score or, time to first severe flare, and mean number of flares and severe flares over 52 weeks and weeks 24

to 52 (Figure 6). Tests for statistical heterogeneity of ORs and HR outcomes were not significant. These results, and those from pooling individual patient data from the two trials prior to logistic regression, mask the systematic difference between trials across all outcomes.

[Insert Figure 6 here]

DISCUSSION

We undertook a systematic review of the clinical effectiveness of belimumab, a new treatment targeted at systemic lupus erythematosus (SLE) patients with anti-nuclear and /or anti-ds DNA autoantibodies. We performed an extensive search and systematic review of both completed and on-going trials using a number of databases and by checking reference lists. Data were extracted independently and studies were quality assessed. Random effects meta-analysis was undertaken.

We identified three RCTs (L02, BLISS-52, BLISS-76) reporting data on over 2000 patients. In contrast to the BLISS trials, L02 recruited patients who were not necessarily current carriers of anti-nuclear or anti ds DNA antibodies at study commencement. L02 failed to demonstrate clinical effectiveness for its primary end points.[26] Meta-analysis of the BLISS studies showed a benefit of belimumab with the main primary outcome (SRI), showing improvement at 52 weeks (OR 1.63; 95% CI: 1.27-2.09 p<0.001) although by week 76, the proportion of SRI responders in the BLISS-76 trial ceased to reach statistical significance (OR 1.31; 95% CI: 0.92–1.87 p=0.1323). There were no significant differences between placebo and intervention groups in quality of life or adverse events.

We found that the benefits of belimumab were systematically greater across the board (although not significantly so) in the BLISS-52 trial and although tests for statistical heterogeneity were negative, geographical location of study centres and the racial background and ethnicity of participants varied considerably. If the two BLISS trials were drawn from the same underlying populations, whilst one might expect outcomes to differ, we would anticipate that this would occur randomly between trials– some better some worse than the other.

A few studies have directly assessed the existence of and importance of geographical differences in trial outcomes.[32-34] Key factors contributing to such differences are variation in underlying patient population characteristics and variation in study execution. Vickers et al,[33]

found that Eastern Asian and Eastern European studies had a higher proportion of positive trial results when compared to other countries. This is seen in the present case for the primary outcome where both the belimumab and placebo response rates in BLISS 52 study were greater than those in BLISS-76 and, remarkably, the placebo response rate in BLISS-52 (49%) was greater than that for the belimumab arm of BLISS-76 (43%). O'Shea and DeMets also report that within the Beta-Blocker Heart Attack Trial (BHAT), not only was there a difference in the direction, but also in the size of treatment effect between Canada and the US, although it should be noted that the original aim of that trial was not investigation of international differences in treatment effect.[35] One study found that 96-99% of the total variance in the "*Global utilisation of strategies to open occluded coronary arteries IV acute coronary syndromes*" (GUSTO IV ACS) trial could be accounted for by patient-level factors.[36]

International trials need to harmonise training of investigators, patient selection, treatment management, thresholds to centre admission, access to facilities, ascertainment of endpoints and, by implication, results of interest [37-44] since it is possible that in centres in different countries these factors may differ systematically.[37] Equally, underlying differences in populations and countries (ethnicity, genetics, socio-economic status and health-care systems), and the nature and epidemiology of SLE according to ethnic background may result in differences in reporting of outcomes and pooled results.

The outcomes used in the BLISS trials would be unfamiliar to most of the investigators and it is possible that criteria may have differed between countries. In particular the Physician Global Assessment (PGA) is an important element of the outcomes measured (see Figure 1). PGA was measured as an outcome in itself, and it is also incorporated in SRI. PGA is of concern because as a global physician assessment of a patient's SLE status, it is subjective. The investigators reported a nearly 10% difference between the BLISS-52 and BLISS-76 studies in estimates of percentage change in PGA score in intervention groups at week 24 compared to baseline and this single result in one of the two trials is likely to have had an important influence on findings of the effectiveness of belimumab in SLE patients.

The latest results of belimumab in patients with SLE (phase II study design, uncontrolled extension study) reported that of 449 patients with active SLE (USA/Canada) 177 (39.4%) patients remained on treatment after 7 years of therapy (i.e. approximately 1746 cumulative patients-years) and that this subgroup exhibited durable sustained improvement in SLE disease activity (SRI and PGA).[30]

CONCLUSIONS

In conclusion, systematic review and random effects meta-analysis of two RCTs of belimumab for patients with autoantibody positive SLE demonstrated positive results in the main outcome at week 52. However, in view of the different populations studied at different locations in the BLISS trials and the consistently superior results from one trial compared to the other, we consider that population heterogeneity, geographical differences and variation in trial conduct and outcome assessment, may have played a role in influencing outcomes. However the generalisability of results pooled meta-analytically or by logistic regression should be viewed with caution and we suggest that it is too early to draw strong conclusions in this case.

ARTICLE FOCUS

- SLE is a complex multi-organ auto-immune disease subject to relapse and remission.
- Patients almost always have fatigue, skin rashes and arthritis and there is a wide variety of other problems which the disease can cause.
- Belimumab is a new treatment specifically targeted against SLE.

KEY MESSAGES

1. Combining the results from two RCTs suggests that belimumab is clinically effective for SLE patients.
2. However, all outcomes were systematically superior in one trial compared with the other.
3. Different trial conduct and populations mean that it is too early to draw generalisable conclusions.

STRENGTHS AND LIMITATIONS

- At first sight combined meta analytic evidence suggests that belimumab is clinically effective for patients with severe SLE.
- We suggest that it is too early to draw strong conclusions because the two relevant trials cover different populations in different countries and there may be differences in trial conduct and outcome assessment.

Acknowledgements

The authors would like to thank the National Institute for Health Research, Health Technology Assessment programme for funding this work.

Funding statement

This work was supported by the National Institute for Health Research, Health Technology Assessment programme [grant number 10/73/01].

Competing interest statement

No conflicts of interest.

Contributions:

N-BK: Conception and design. Data analysis and interpretation. Drafting the article. Critical revisions for important intellectual content. Approval of final article for submission.

MC: Conception and design. Data analysis and interpretation. Literature review. Interpretation of results. Drafting the article. Critical revisions for important intellectual content. Approval of final article for submission.

AG: Interpretation of results. Critical revisions for important intellectual content.

PS: Literature review. Interpretation of results. Critical revisions for important intellectual content.

SM: Data analysis and interpretation. Interpretation of results. Critical revisions for important intellectual content.

LH: Literature review. Interpretation of results. Critical revisions for important intellectual content.

RC: Literature review. Critical revisions for important intellectual content.

EC: Interpretation of results. Critical revisions for important intellectual content.

CG: Interpretation of results. Critical revisions for important intellectual content.

AC: Conception and design. Interpretation of results. Drafting the article. Critical revisions for important intellectual content. Approval of final article for submission.

All authors read and approved the final manuscript.

Reference List

(1) Bernatsky S, Joseph L, Pineau CA, Tamblyn R, Feldman DE, Clarke AE. A population-based assessment of systemic lupus erythematosus incidence and prevalence--results and implications of using administrative data for epidemiological studies. *Rheumatology (Oxford)* 2007; 46(12):1814-1818.

(2) Pons-Estel GJ, Alarcon GS, Scofield L, Reinlib L, Cooper GS. Understanding the epidemiology and progression of systemic lupus erythematosus. *Semin Arthritis Rheum* 2010; 39(4):257-268.

(3) Gordon C, Li CK, Isenberg DA. Systemic lupus erythematosus. *Medicine* 2010; 38(2):73-80.

(4) Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1997; 40(9):1725.

(5) Tan EM, Cohen AS, Fries JF, Masi AT, McShane DJ, Rothfield NF et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1982; 25(11):1271-1277.

(6) Griffiths B, Mosca M, Gordon C. Assessment of patients with systemic lupus erythematosus and the use of lupus disease activity indices. *Best Practice and Research in Clinical Rheumatology* 2005; 19(5 SPEC. ISS.):685-708.

(7) Somers EC, Thomas SL, Smeeth L, Schoonen WM, Hall AJ. Incidence of systemic lupus erythematosus in the United Kingdom, 1990-1999. *Arthritis Rheum* 2007; 57(4):612-618.

(8) Pons-Estel GJ, Alarcon GS, McGwin G, Jr., Danila MI, Zhang J, Bastian HM et al. Protective effect of hydroxychloroquine on renal damage in patients with lupus nephritis: LXV, data from a multiethnic US cohort. *Arthritis Rheum* 2009; 61(6):830-839.

(9) Pons-Estel GJ, Gonzalez LA, Zhang J, Burgos PI, Reveille JD, Vila LM et al. Predictors of cardiovascular damage in patients with systemic lupus erythematosus: data from LUMINA (LXVIII), a multiethnic US cohort. *Rheumatology (Oxford)* 2009; 48(7):817-822.

(10) Bernatsky S, Boivin JF, Joseph L, Manzi S, Ginzler E, Gladman DD et al. Mortality in systemic lupus erythematosus. *Arthritis and Rheumatism* 2006; 54(8):2550-2557.

(11) Kurien BT, Scofield RH. Autoantibody determination in the diagnosis of systemic lupus erythematosus. *Scand J Immunol* 2006; 64(3):227-235.

(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; 62(1):222-233.

(13) Summary of product characteristics. 2012 Available from: URL: ema.europa.eu/Find medicine/Human medicines/European Public Assessment Reports Accessed January 2012

(14) American College of Rheumatology. Belimumab for Systemic Lupus Erythematosus. 2011 Available from: URL: http://www.rheumatology.org/publications/hotline/2011_03_15_belimumab.asp Accessed January 2012.

(15) New Drug Bulletin. Belimumab (Benlysta - Human Genome Sciences, Inc). 2012 Available from: URL:http://healthcare.utah.edu/pharmacy/bulletins/NDB_228.pdf Accessed January 2012.

(16) Isenberg DA, Rahman A, Allen E, Farewell V, Akil M, Bruce IN et al. BILAG 2004. Development and initial validation of an updated version of the British Isles Lupus Assessment Group's disease activity index for patients with systemic lupus erythematosus. *Rheumatology* 2005; 44(7):902-906.

(17) Gordon C, Bertsias G, Ioannidis JPA, Boletis J, Bombardieri S, Cervera R et al. EULAR points to consider for conducting clinical trials in Systemic lupus erythematosus. *Ann Rheum Dis* 2009; 68(4):470-476.

(18) Burgos PI, McGwin G, Jr., Pons-Estel GJ, Reveille JD, Alarcon GS, Vila LM. US patients of Hispanic and African ancestry develop lupus nephritis early in the disease course: data from LUMINA, a multiethnic US cohort (LUMINA LXXIV). *Ann Rheum Dis* 2011; 70(2):393-394.

(19) Khan KS, Ter Riet G, Glanville J, Snowdon AJ, Kleijnen J. Reviews of Research on Effectiveness. CRD's Guidance for Carrying Out or Commissioning Reviews. 2 nd Edition. (ISBN 1900640201) CRD Report No. 4. York: NHS Centre for Reviews and Dissemination (CRD), University of York. 2001 Available from: URL:http://www.york.ac.uk/inst/crd/report4.htm; Accessed January 2012.

(20) Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327(7414):557-560.

(21) Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. 2012 Available from: URL:www.cochrane-handbook.org Accessed January 2012.

(22) DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; 7(3):177-188.

(23) STATA Statistical analysis and software. 2011 Available from: URL:http://www.stata.com/ Accessed April 2012.

(24) 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(9767):721-731.

(25) Furie R, Petri M, Zamani O, Cervera R, Wallace DJ, Tegzova D et al. A phase III, randomized, placebo-controlled study of belimumab, a monoclonal antibody that inhibits B lymphocyte stimulator, in patients with systemic lupus erythematosus. *Arthritis Rheum* 2011; 63(12):3918-3930.

(26) 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(9):1168-1178.

(27) GlaxoSmithKline. GSK1550188 A 52 week study of belimumab versus placebo in the treatment of subjects with SLE located in North East Asia. 2011 Available from: URL:http://clinicaltrials.gov/ct2/show/NCT01345253 Accessed January 2012.

(28) 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. 2010 Available from: URL:http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/ArthritisDrugsAdvisoryCommittee/UCM233581.pdf Accessed January 2012.

(29) U.S.Food and Drugs Administration. FDA briefing package. 2010 Available from: URL:http://www.fda.gov/downloads/advisorycommittees/committeesmeetingmaterials/drugs/arthritisdrugsadvisorycommittee/ucm233579.pdf Accessed January 2012.

(30) GlaxoSmithKline. Single Technology Appraisal (STA): Belimumab for the treatment of active autoantibody-positive systemic lupus erythematosus. 2011 Available from: URL:http://www.guidance.nice.org.uk/TA/live/13307/56490/56490.pdf Accessed January 2012.

(31) Egger M, Smith GD, Altman DG. Systematic Reviews in Health Care, Meta-analysis in context. 2 ed. BMJ Books; 2001.

(32) Pan Z, Trikalinos TA, Kavvoura FK, Lau J, Ioannidis JP. Local literature bias in genetic epidemiology: an empirical evaluation of the Chinese literature. *PLoS Med* 2005; 2(12):e334.

(33) Vickers A, Goyal N, Harland R, Rees R. Do certain countries produce only positive results? A systematic review of controlled trials. *Control Clin Trials* 1998; 19(2):159-166.

(34) Zhang D, Freemantle N, Cheng KK. Are randomized trials conducted in China or India biased? A comparative empirical analysis. *J Clin Epidemiol* 2011; 64(1):90-95.

(35) O'Shea JC, DeMets DL. Statistical issues relating to international differences in clinical trials. *Am Heart J* 2001; 142(1):21-28.

(36) Chang WC, Midodzi WK, Westerhout CM, Boersma E, Cooper J, Barnathan ES et al. Are international differences in the outcomes of acute coronary syndromes apparent or real? A multilevel analysis. *J Epidemiol Community Health* 2005; 59(5):427-433.

(37) White HD. International differences: selection, noise, or real? *Eur Heart J* 2000; 21(5):339-342.

(38) Fox KA, Goodman S, Bigonzi F, Le L, V, Cohen M. Inter-regional differences and outcome in unstable angina; analysis of the international ESSENCE trial. Efficacy and Safety of Subcutaneous Enoxaparin in Non-Q-wave Coronary Events. *Eur Heart J* 2000; 21(17):1433-1439.

(39) Sinnaeve PR, Van de Werf FJ. Global patterns of health care for acute coronary syndromes. *Curr Opin Cardiol* 2004; 19(6):625-630.

(40) Faergeman O, Kjekshus J, Cook T, Pyorala K, Wilhelmsen L, Thorgeirsson G et al. Differences in the treatment of coronary heart disease between countries as revealed in the Scandinavian Simvastatin Survival Study (4S). *Eur Heart J* 1998; 19(10):1531-1537.

(41) Berger JS, Stebbins A, Granger CB, Ohman EM, Armstrong PW, Van de Werf F et al. Initial aspirin dose and outcome among ST-elevation myocardial infarction patients treated with fibrinolytic therapy. *Circulation* 2008; 117(2):192-199.

(42) Welsh RC, Chang W, Goldstein P, Adgey J, Granger CB, Verheugt FW et al. Time to treatment and the impact of a physician on prehospital management of acute ST elevation myocardial infarction: insights from the ASSENT-3 PLUS trial. *Heart* 2005; 91(11):1400-1406.

(43) Yusuf S, Flather M, Pogue J, Hunt D, Varigos J, Piegas L et al. Variations between countries in invasive cardiac procedures and outcomes in patients with suspected unstable angina or myocardial infarction without initial ST elevation. OASIS (Organisation to Assess Strategies for Ischaemic Syndromes) Registry Investigators. *Lancet* 1998; 352(9127):507-514.

(44) Bleyer AJ, Hylander B, Sudo H, Nomoto Y, de la Torre E, Chen RA et al. An international study of patient compliance with hemodialysis. *JAMA* 1999; 281(13):1211-1213.

June 4, 2013

Mr. Richard Sands
Managing Editor, BMJ Open

RE: Manuscript ID bmjopen-2013-002852
Title: Belimumab: a technological advance for SLE patients? Report of a systematic review and meta-analysis

Dear Mr Richard Sands,

Please find enclosed our revised manuscript, which addresses the reviewers' concerns and suggestions. What follows is a point-by-point response to the comments provided as part of the review process. Each group of responses has been numbered to correspond with those on the comments. Moreover, in the revised manuscript we have highlighted in red colour the areas that have been substantively modified compared to the original submission.

We would like to thank the reviewers and managing editor for thoughtful comments and suggestions. We truly appreciate your interest in our work. We believe that as a result of the review process our paper has greatly improved and hope that it is now acceptable for publication in BMJ Open.

Editor:

From the managing editor:
None of the references include dates. Please add where possible.

Reply: Thank you for the comment. We have now included dates of references as suggested.

Reviewer 1:
Reviewer: Peter Watson
Statistician

MRC Cognition and Brain Sciences Unit
15 Chaucer Road
Cambridge
UK
CB2 7EF

I have no conflicting interests with the research presented in this study.

1. There appear to be many analyses and response variables without any particular one being of primary interest. I have a concern given the heterogeneity of the ethnicity (page 10 second paragraph) and the small implied number of studies (page 3, results, first sentence) of generalisability of the results and representativeness to other populations. The degree of between study heterogeneity could be stated using I^2 and, if not already, accounted for in deriving pooled

estimates. Other aspects of the results and figures could be described in greater depth (see comments below) including labeling and captioning of all the figures and more clearly linking the results in the text to those in the figures and stating which analyses are used to produce the results plotted in the figures.

Reply: We would like to thank the reviewer for his comments on our manuscript. We have now carefully checked the manuscript to account for these comments and suggestions.

The reviewer is right, although the main primary outcome to determine the effectiveness of belimumab was the Responder Index (SRI) at week 52, we also examined other outcomemeasures for the three RCTs that evaluated belimumab effectiveness e.g. examining the SLE Responder Index (SRI) at week 76. We also included those outcomes identified by the belimumab investigators in their protocol as "major secondery and other outcomes". We have now clearly identified the primary outcome designated in the RCTs (namely SRI at 52 weeks), we have stated that this is also our primary outcome, and have included text to explain the origin of this novel outcome measure as developed between the FDA and the belimumab trialists.

We have attempted to highlight more explicitly that our manuscript concerns the generalizability of pooled results and that these should be viewed with caution. We noted that population heterogeneity; geography and / or variation in trial conduct may be influence results; we have removed reference to "*hidden confounders*". Although formal tests for statistical heterogeneity were negative, BLISS-52 results were systematically more favourable for all measured outcomes.

These elaborations on the interpretation of our results are found mainly in lines:

89-95; 197-201; 261-271.

2. A couple of references on meta-analysis that may be of use I put in the comments below which may be worth adding to the bibliography.
Reply: We have added the Higgins reference as suggested; the reference for publication bias has not been added because it was not possible to ascertain if there was publication bias with only two RCTs; we have added text to this effect (line 154) the reference to the Cochrane Handbook (number 21) was therefore considered sufficient. Ref 21 (page 317) recommends at least 10 studies would be required for analysis of small study bias and we have been guided by this.

3. It is not clear how the unnumbered and uncaptioned figures (pages 27-29) relate to the results (pages 8-9) and if, and how adjusted, the pooled odds ratios quoted on page 8 (first paragraph lines 8-9) relate to the binary responses in Table 3 (page 32). I think the lack of statistical significance of both inter-study heterogeneity of effect sizes (page 10 first paragraph) and the confidence intervals of the figures mostly containing values ('1' for odds ratios and '0' for mean differences) suggesting no group differences could be down to limited power and possibly low sample sizes. This weakness may be mitigated by the number of point estimates suggesting a (hopefully clinically meaningful) benefit of the belimumab treatment but this need to be motivated in the text.

Reply: We have revised and numbered captions of figures and relate them clearly to the results section as suggested. Lines 214-218 explain how the results depicted in figure were derived; lines 227-230 explain how the adjusted odds ratios were derived / reported. As for the weakness of the study as mentioned by the reviewer, the reviewer makes an important point. The reviewer

indicates that confidence intervals “suggesting no group differences” might be attributable to lack of power is of course probable, however the modest effect size (small benefit of belimumab) is a major contributory factor. Due to the scarcity of RCTs evaluating the effectiveness of belimumab, we restricted our study to the available evidence. The three RCTs combined investigated 2133 SLE patients, which may be a good sample size for this type of rare condition (e.g. the SLE Rituximab trial, the only other major recent trial for SLE, recruited 184 patients into two arms).

4. There are limitations (conclusions section on page 3) concerning 'hidden confounders' and interpretability of pooled estimates. It is not clear to me (see later comments) if this is 'merely' downplaying a pooled estimate and usefulness of a meta-analysis as the (limited number, three, of) populations being pooled are so different from each other or, more seriously, if there could be possible uncontrolled differences in clinically meaningful characteristics (confounders) between the placebo and treatment groups in one or more studies which would render any differences between the groups problematic to interpret as they could be simply due to factors other than the belimumab treatment. There is some mention of stratified randomisation on page 7 (start of second paragraph) but no details of what factors were used as stratifers.

Reply: We have attempted to clarify these issues. We have removed the phrase “hidden confounders” and have explicitly considered the influence of geographical / ethnic / trial conduct differences between the BLISS trials by first pointing to the systematic difference in results between B52 and B76 (lines: 236-242; 262-272) and by alluding to the ethnic / geographical data presented in Table2 and Figure3; lines 262-272) . We now provide explicit information about the stratification undertaken in the BLISS trials and the use of strata in adjusting results reported in the published accounts (lines 182-184; 227-229). The limitation we mentioned is not only limited to the general applicability of the nature of meta-analysis but also to real limitations due to confounders such as the geographic location and the ethnicity where the studies were conducted.

5.This study compares a group using a new treatment for multi-organ auto-immune disease, Belimumab, with a placebo group by, on pages 8 and 9, obtaining confidence intervals for odds ratios (for a series of binary responses) plotted in the figures on page 27 and mean differences (for continuous ones) plotted in the figure on page 28 and reports a meta-analysis on page 9 for each of five (?) outcomes which look at the group effect which I suspect may be plotted in the figures on page 29.

Reply: Please see the method section of the paper. We performed a meta-analysis of two randomized controlled trials (RCTs) of belimumab against placebo or best supportive care. To improve clarity we have edited the figure captions and Method sections. The Meta-analysis figure (Figure 6) shows the results of random effects meta- analysis of the two BLISS trials for each of 14 outcomes designated by belimumab trialists as primary or major secondary or “other major” outcomes. For convenience of viewing we combined the results for different types of outcome into a single figure (binary, time to event and continuous) using Excel.

6. I, unfortunately, found the description (on pages 8 and 9) and presentation of the results (in the figures on pages 27-29) confusing and imprecise making it difficult to marry together the description of the results in the text and the confidence intervals plotted in the figures. The structure of the data being analysed needs to be fleshed out in the body of the text to help understanding of the results e.g. I am not sure if the meta-analyses are pooling across different

studies or different subgroups within a single study or precisely what the SLE in the title of this paper stands for (it presumably is an abbreviation?)

Reply: We have defined SLE in the title and text. We have clarified the results and figures presented to explain that the pooling was across different studies (the two RCTs). We also present within-study results for the primary outcome according to different geographical subgroups (lines 262-272). With many outcomes and sub-groups analysis it became difficult for the reader we consider that we have improved the paper in this regard.

7. In particular the figures on pages 27-29 were not numbered or captioned which made it more difficult to know which analyses and effect sizes (odds ratio or 'mean difference') they were referring to and, in particular, which is the Figure 6 listed as corresponding to the meta-analyses reported briefly in the second paragraph on page 9. There is also an effect size called the 'hazard ratio' in a figure on page 27 which does not seem to be defined in the text. There are a lot of responses (listed both within the figures and represented by these different figures on pages 27-29 and also mentioned as a basis for various meta-analyses in the first sentence of the second paragraph on page 9). It is not clear to me if the results of analyses of these separate responses are being presented or discussed separately or together.

Reply: Thank you for these comments. We have now explained the derivation of the hazard ratio results (lines 236-242). We have attempted to explain why so many outcome measures exist for SLE (lines 88 to 95) and how this led to the development of the SRI measure. These results are mainly, but not exclusively, discussed together since the most noticeable feature common to all is the better performance of belimumab in B52 relative to B76 (lines 262 265; 307-313).

8. Page 6. I think it makes more sense grammatically to say at the end of the first sentence of the 'Statistical analysis' paragraph on page 6 that odds ratios and mean differences 'were calculated for binary and continuous outcomes respectively'. Two reviewers are mentioned on page 6 under 'inclusion criteria' as assessing inclusion of studies. Was this assessment done independently by the two raters and, if so, could a kappa statistic, or alternative, be quoted to show inter-rater agreement?

Reply: We have added modified the sentence as suggested and clarified the independence and tasks of the two reviewers (lines 138-145).

9. Pages 6 and 8. The statistical analysis on page 6 mentions 'unadjusted odds ratios'. Adjusted odds ratios are then presented (fifth line from bottom of first paragraph on page 8) but it doesn't mention in either sentence what these odds ratios are adjusted for or how or why both unadjusted and adjusted odds ratios are used. If its ok to use unadjusted odds ratios why adjust them?

Reply: We have now clarified the use of adjusted and unadjusted odds ratio to make it clear to the reader why both were presented (lines 214-218; 227-229)

10. Pages 6 and 8. Is the 'mean difference' reported in the 'Statistical analysis' paragraph on page 6 and in the second paragraph on page 8 a standardised group one such as Cohen's d if

you are wishing to compare results for different responses which may have different scales?

Reply: The mean difference' reported in the 'Statistical analysis' in paragraph 6 and 8 is mean difference' reported in the BLISS RCTs. Each outcome used the same assessment tool in both trials and "standardized mean difference" such as Cohen's d was not appropriate.

11. Pages 6 and 9. I would like to see in the meta-analysis (second paragraph on page 9) the value of I^2 and any associated p-value, which was used (as stated on page 6 in the second last paragraph labelled 'statistical analysis') to test for the heterogeneity of effect size as this is an important test given that the degree of study heterogeneity is referred to throughout this paper. There are rules of thumb for small, medium and large values of I^2 that could be used. A value of 0% indicates no observed heterogeneity, 25%-49% is low heterogeneity, 50%-74% is moderate and 75% and above is large (Higgins et al, 2003). You could also mention in the statistical analysis paragraph on page 6 if you used a Der Simonian pooled estimate for the effect sizes or a fixed effect one such as the Mantel-Haenszel estimate for odds ratios in the meta-analysis as you found (page 9) little or no between study variation.

Reply: We now explain that we used the random effects method of DerSimonian Laird (line 156) to pool effect sizes. We anticipated heterogeneity so a random effects model was more appropriate in this case than the fixed effects model. We have now displayed (in Figure 6) the value of I^2 and the associated p-value as suggested, and we have tightened the text so the lack of statistical heterogeneity refers specifically to binary and time to event outcomes.

12. Page 7. Second paragraph, line 1 mentions 'stratified randomisation' was used. What factors were stratified for and for what factors were the arms 'well balanced'?

Reply: We have now clarified this (lines 227-229). Baseline balance included values for: proteinuria, disease duration, gender, race, IgG, autoantibody, and complement levels , baseline SLEDAI and PGA scores, BILAG organ domain involvement and SLICC Damage Index score; we have now included this information in the caption to figure 5.

13. Page 8. I am not clear from the results on pages 8 and 9 how we should go about interpreting the confidence intervals in the figures on pages 27 to 29. Confidence intervals for odds ratios 'pooled across trials' are presented in the first paragraph (line 6) on page 8 but these are not graphed in the figures on pages 27 and 28 and I am not sure how these tie in with the confidence intervals in the figures. Are the results on lines 8-9 of the first paragraph on page 8 pooling odds ratios across all the binary variables mentioned in Table 3 (page 32) in BLISS-52 and BLISS-76 and is a pooled odds ratio interpretable when pooling over apparently different tests? The 'pooled across trials' implies some meta-analysis may have been performed to yield these pooled odds ratios.

Reply: In the figures referred to (on pages 27 to 29) ORs are unadjusted (now explained more clearly lines 214-218). Additionally we have explained that the BLISS trial journal articles and the manufacturer's submissions to the FDA and to NICE used a logistic regression model (individually for each trial in the journal articles, and after pooling populations in the case of the

submissions to the approval authorities; lines -229; 227 and 276-279).

14. Page 9. The first line of the second paragraph on page 9 implies that a meta-analysis is performed on each of at least five different responses (as meta-analyses usually pool over trials measuring effect sizes using the same response and groups) and there is a mention of figure 6 which is the last figure in the paper presumably the one on page 29 yet I can't see six separate plots here. I would also have expected to see a confidence interval for a pooled effect size at the base of each of the forest plots corresponding to the meta-analysis of each response.

Reply: The text referring to Figure 6 has been clarified (lines 282-285). The figure has been redrawn and figure caption improved to correct for errors and improve clarity for the reader.

15. Page 9. The last sentence of the first paragraph on page 9 mentions there were various causes of death but does not mention what these were which I would have thought would be of interest in giving a background to the data. I am not sure if the 'study level' referred to in the first sentence of the second paragraph on page 9 refers to separate subgroups within studies or, the usual pooling unit of pooling in meta-analyses, separate studies.

Reply: We have now included the causes of death (lines 250-251). The term "study level" was used to distinguish the results presented from those in the manufacturer's submission to the FDA in which IPD from the two BLISS trials was pooled prior to logistic regression analysis; hopefully this is now clear from the text (lines 281-287)

16. Pages 9 and 29. I don't see any mention of a funnel plot to test and adjust for any possible publication bias. This analysis, at least, is usually performed and plotted routinely in meta-analyses including those submitted to this journal. Other tests can also be used – see, for example, Peters et al. (2010).

Reply: The reviewer makes a potentially important point here. We did not include a formal test of small study bias because there are too few trials evaluating the effectiveness of Belimumab (see line 154 with accompanying reference 21) . We believe, a test of publication bias in this context may not be useful.

17. Pages 9, 27-29. Page 9 implies a meta-analysis has been performed and, in light of this, I was surprised to see the size of the point estimates in the middle of all the confidence intervals plotted in the figures on pages 27-29 looking the same size as these usually differ in size as they are proportional to the weighting given to the studies in the meta-analysis to construct a pooled estimate. I also think, therefore, for the forest plot(s) you could add in a column by the plot showing the value of the weights used to confirm the studies had a similar weighting used in constructing the pooled estimate.

Reply: All points are the same size because each refers to the pooled estimate for a particular outcome, not to a single study estimate given a specific weight in the analysis. We hope the text and figure caption and redrawn figure now make this clearer.

18. Page 10. The first paragraph mentions that there was no heterogeneity found (across the studies or subgroups?) in the BLISS-52 trial but, counterintuitively, the racial background and ethnicity of participants 'varied considerably' and concludes there should be heterogeneity which confuses the conclusion and makes one start to doubt the tests of heterogeneity that have been used in this analysis as basis for obtaining pooled estimates. I am not sure if the conclusion (page 10 first line of first paragraph) that the benefits of belimumab are 'greater across the board' is warranted looking at the confidence interval plots on pages 27-29 since most of these intervals contain either an odds ratio of one or a zero difference which both correspond to no difference. One might possibly argue that, ignoring variances, the bulk of the point estimates, comprising odds ratios and mean group differences, are benefitting the use of the treatment, belimumab, but this needs to be carefully argued in the light that few of them are statistically significant and given the acknowledged heterogeneity (on page 10) which the authors may wish to account for if they have not done so already in obtaining pooled effect sizes despite the 'usual' tests of these not flagging this which may be due to lack of power from heterogeneity across only three studies being tested.

Reply: We performed the I squared test for statistical heterogeneity between the two BLISS trials used in the meta analyses and found low values for all outcomes. But we believe that there are other sources of heterogeneity (geographical, trial conduct etc.) which have exerted a systematic influence on the outcomes, the major indicator of this influence being the consistently superior performance of one trial compared to the other across multiple outcomes. Hopefully the new text (e.g. lines 262 -272) explains this more clearly. The fact that BLISS 76 outcomes almost always fail to reach statistical significance is now brought out more clearly (e.g. lines 253-260); even though the fact that the primary outcome in BLISS 76 was satisfied on extending observation to 76 weeks eliminates the statistical significance of the SRI. While the lack of statistical significance may be attributable to some extent to lack of power it is also clear that effect sizes in BLISS 76 are modest.

19. Pages 27, 28 and 29. The figure(s) containing the forest plots need to be numbered and captioned. Is it necessary to both plot and quote the confidence intervals for group differences in these figures? Would simply plotting these confidence intervals be enough?

Reply: We have now numbered and captioned the plots.

Figures 3 and 4 present the comparison (intervention versus control) separately for the two BLISS trials because this highlights the fact that BLISS 52 always gives a better result for belimumab than does BLISS 76. We think the CIs are necessary because again they highlight the difference between the trials.

20. The plot on page 28 plots hazard ratios (as opposed to rates?) in the 'time to event' figure which are, generally, not the same as odds ratios. The hazard ratios should be defined in the text but I can't see any mention of hazard ratios anywhere else in the paper (e.g. in the statistical analysis paragraph on page 6 or in the results sections on pages 8 and 9).

Reply: Thank you for the comments. We have now explained the hazard ratios in lines 236-242.

21. The study does not explicitly state on page 9 in the meta-analysis results section how many

trials are being pooled to obtain pooled effect sizes in the meta-analyses although elsewhere (for example on page 3, first line in first paragraph) three trials are mentioned and two 'relevant trials' (page 2 second bullet point under 'strengths and limitations'). Usually one has sufficient numbers of studies being pooled to make any results generalizable across different types of study to different populations. I mention this, as three trials, if this is the number used, does not seem very many for a meta-analysis particularly one where there is considerable between study heterogeneity at least in ethnicity (as already noted in the first paragraph on page 10), and as some of the plots in the figures on pages 27-29 only contain four rows (and then assuming one would be pooling BLISS-52 and BLISS-76 whose pooling might be questionable given separate confidence intervals are presented for these in the fifth last row from the end of the first paragraph on page 8).

Reply: We performed the meta-analyses using outcomes from two randomized controlled trials (the two BLISS trials). (This is now more explicit in lines 281-282, and in the caption to the figure 6). We explain why the L02 trial was only used in assessing safety outcomes on lines 194-198. Problems in interpreting what is represented in the figures have been addressed in figure captions and with more explicit description in the body of the text.(lines 212-218 and 237-243).

22. On page 3 (in the conclusions paragraph) the fourth line states generalizability of 'pooled results should be viewed with caution' and lines 5 and 6 mention possible 'hidden confounders'. Is this saying that the pooled studies may have differed from one another in many respects (confounders) and/or is it saying there are so many possibly uncontrolled confounders of clinical relevance in these group comparisons that we are looking at group differences (the belimumab treatment group vs the placebo group) that could be due to other clinically meaningful confounding factors which differ between the treatment and placebo groups? The latter could be a serious drawback to interpretability of any results whereas the former would, at least, preclude an interpretable pooled estimate since we would be averaging over such disparate (and few) populations which rather undermines the usefulness of a meta-analysis.

Reply: Perhaps, it was not clear in the previous version of the paper. We have removed reference to "hidden confounders" and have clarified the conclusion section (especially lines 361-365). Please also see our reply in point 4 above.

References

Higgins, JP, Thompson SG, Deeks JJ and Altman DG (2003). Measuring inconsistency in meta-analyses. BMJ, 327, 557-560.

Peters, J.L., Sutton, A.J., Jones, D.R. and Abrams K.R. (2010). Assessing publication bias in meta-analyses in the presence of between-study heterogeneity. Journal of the Royal Statistical Society A, 173(3), 575-591 There is an on-line copy of this paper at http://onlinelibrary.wiley.com/doi/10.1111/j.1467-985X.2009.00629.x/full.

Reviewer 2

Ricard Cervera, MD, PhD, FRCP

Head, Department of Autoimmune Diseases
Hospital Clínic
Barcelona, Catalonia, Spain

Statement: I have no competing interests with the authors of this manuscript

Reviewer: This is an interesting systematic review and meta-analysis of the randomized controlled trials of belimumab in patients with systemic lupus erythematosus. The study was well designed, the results are interesting and the manuscript is well written with well balanced discussion.

Reply: We would like to thank the reviewer for his kind comments on our manuscript.

Once more, we would like to thank the reviewers for thoughtful comments and suggestions. We truly appreciate your interest in our work. We believe that as a result of the review process our paper has greatly improved and hope that it is now acceptable for publication in BMJ Open.

Yours sincerely,

Ngianga-Bakwin Kandala, PhD

FIGURE 1: Summary of the major clinical measures used in SLE trials

SELENA-SLEDAI: Encompasses 24 weighted items scored dichotomously as present or absent in the previous 10 days, thus improvement or worsening of a manifestation is not captured. Overall disease activity is scored over a range of 0 to 105 points. A minimum clinically meaningful score change = a decrease of 6 points (overall improvement) or an increase of 8 points (overall worsening). A designated change in score (≥ 4 points) between baseline and follow up can be used to dichotomise patients into responders or non-responders for overall disease.

BILAG: Includes 86 items grouped in eight 8 organ systems to assesses organ system involvement over the last four weeks compared to preceding four weeks based on physicians intention to treat using classifications ranging from A to E as follows: A = worsening usually requiring intensification of steroids or immunosuppressant treatments; B = worsening usually requiring antimalarials, nonsteroidal anti-inflammatory drugs (NSAIDs), or low dose steroids; C = stable disease (symptomatic therapy); D = improvement; E = system never involved. Unlike SELENA-SLEDAI it can detect worsening or improvement in individual organ system involvement.

PGA: Is employed to monitor change in patient overall disease activity; typically a visual analogue scale is used ranging between no disease = 0, mild disease = 1, moderate disease = 2, and severe disease = 3.

SRI: A composite instrument (combining elements of SELENA-SLEDAI, BILAG and PGA) developed by belimumab-trialists in conjunction with the US FDA. It allows patients to be dichotomised into responders or non-responders according to predefined assessment criteria in each of the component elements, such as: a SELENA-SLEDAI improvement of ≥ 4 points, plus no worsening in PGA score by > 0.3 points, plus no new BILAG organ system involvement scoring category A in one system or category B in two or more systems. An advantage of SRI, over any one of its components used alone, may be that it can detect SLE improvement in some initial manifestation(s) while guarding against the possibility that worsening in organ systems or overall disease activity might be masked.

FIGURE 2: PRISMA 2009 Flow Diagram for Belimumab in SLE RCTs and on-going trials

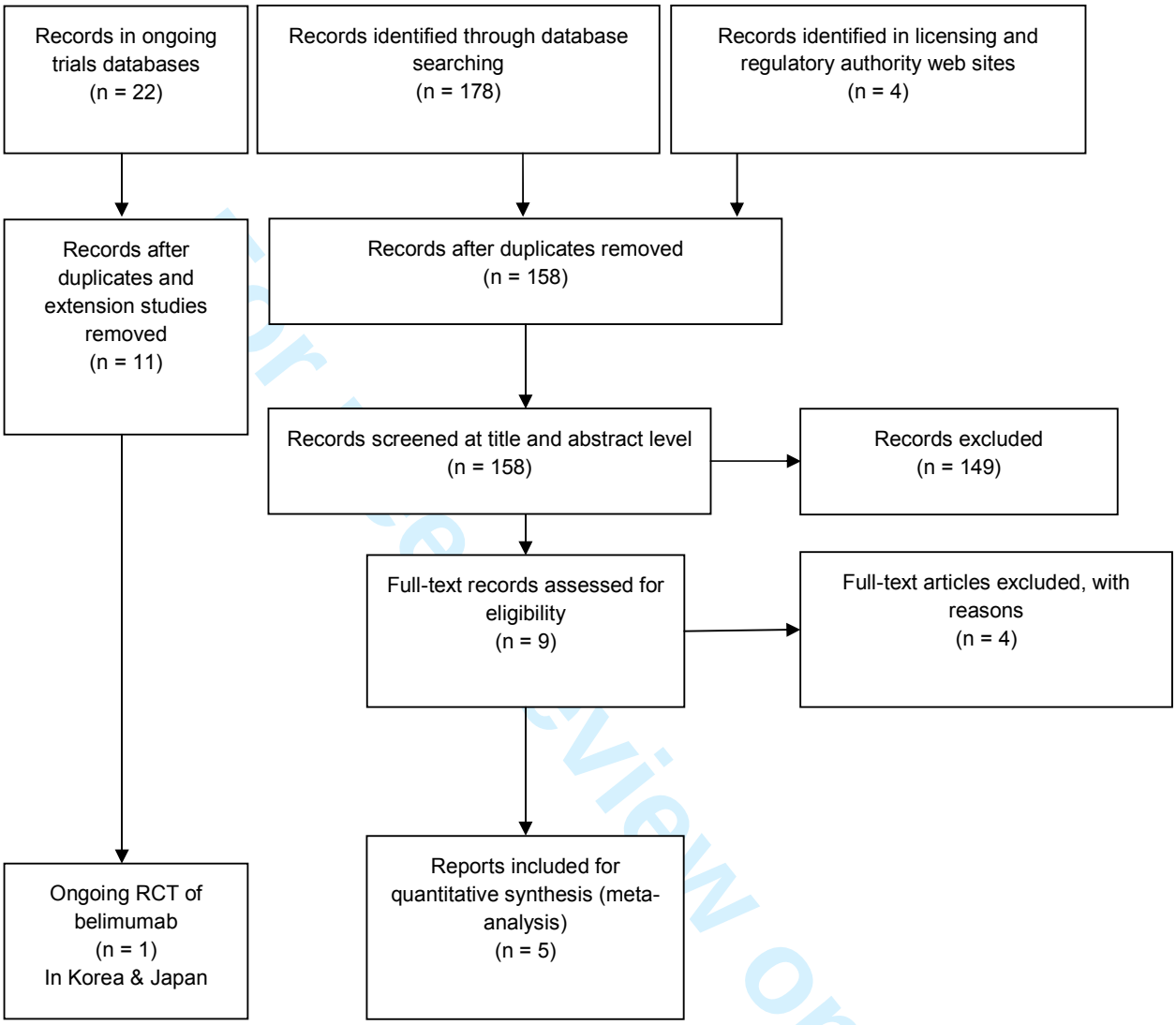


FIGURE 3: Differing centre locations in the BLISS 52 and BLISS 76 multicentre trials

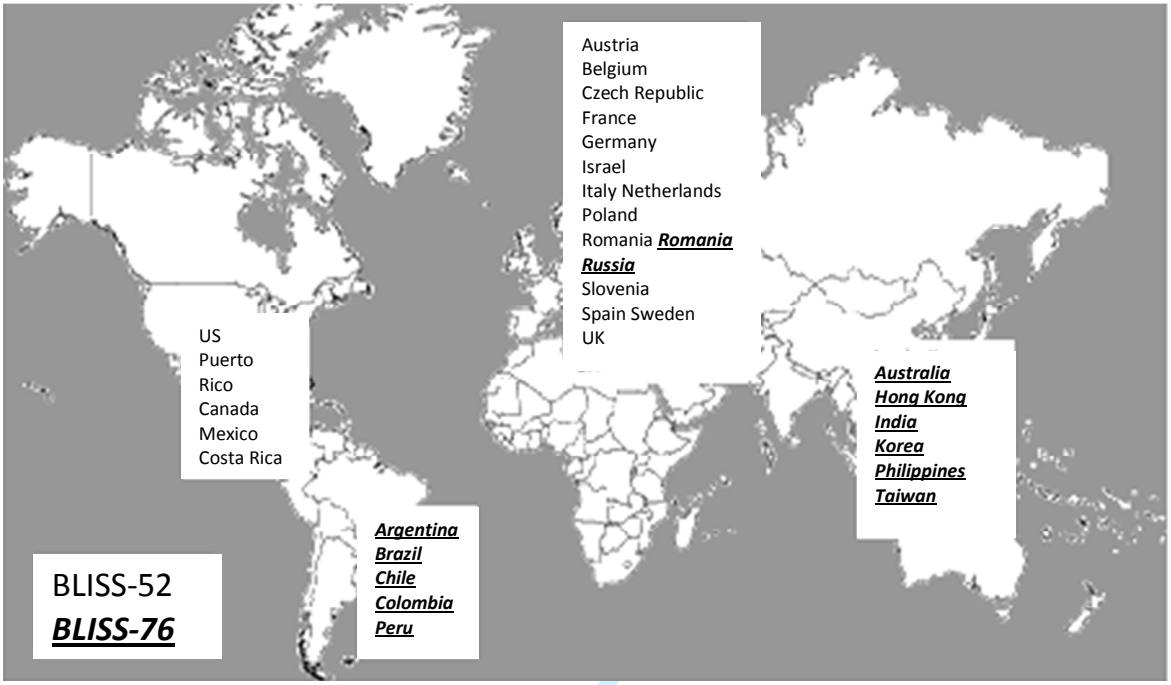


FIGURE 4: Summary of results for major binary and time to event outcomes in belimumab RCTs

Except for safety outcomes the results shown are for the BLISS 52 and BLISS 76 trials. Odds ratios (OR) were calculated from the event rates reported in journal publications; hazard ratios are from data presented in the manufacturer's submission to the FDA. The BLISS trials were well balanced for baseline characteristics (disease, duration, Gender, race, baseline IgG, autoantibody, and complement levels, baseline SLEDAI and PGA scores, BILAG, organ domain involvement, SLICC Damage Index score, and Proteinuria). Safety outcomes are based on data presented in FDA documents.

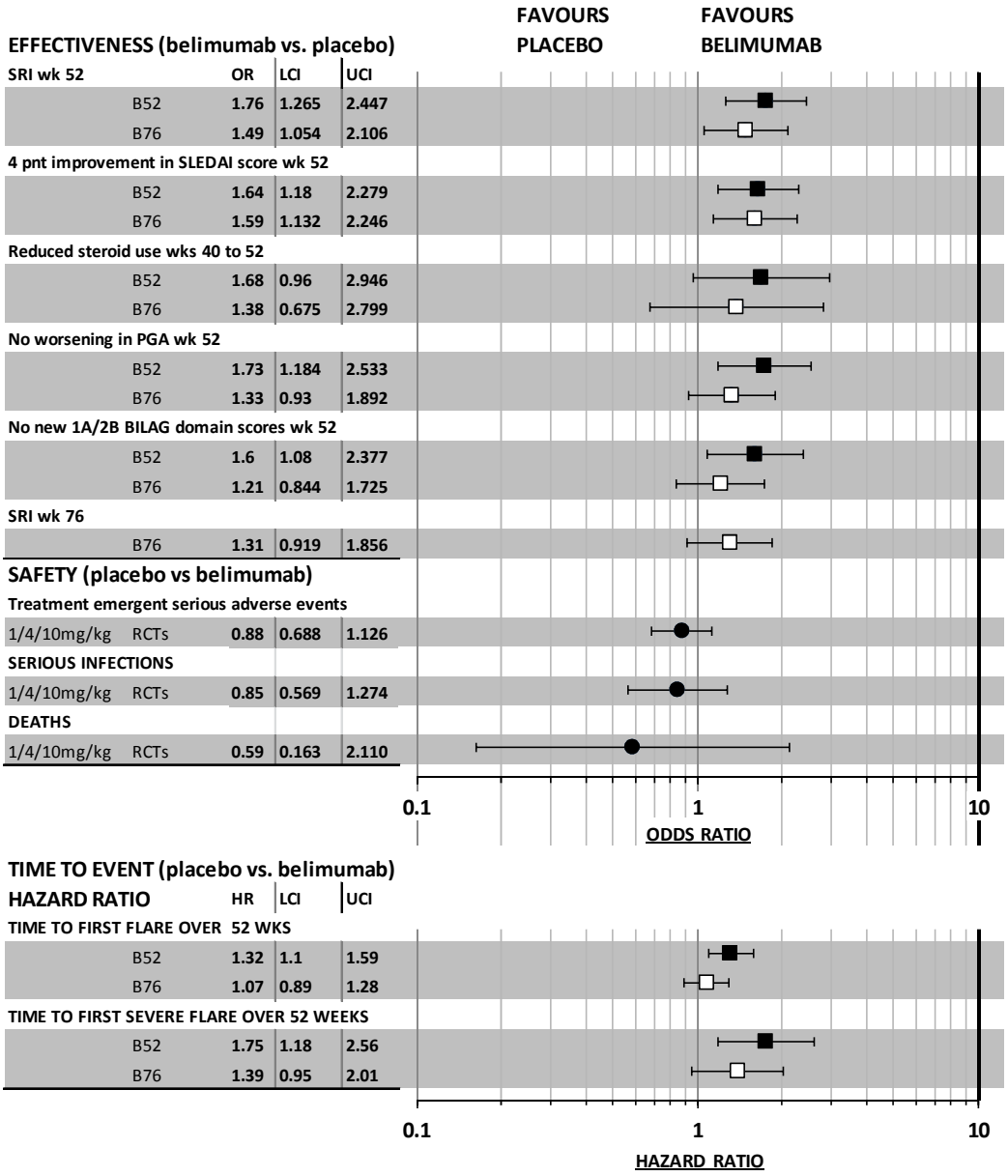


FIGURE 5: Summary of results for major continuous outcomes in BLISS 52 and BLISS 76 trials

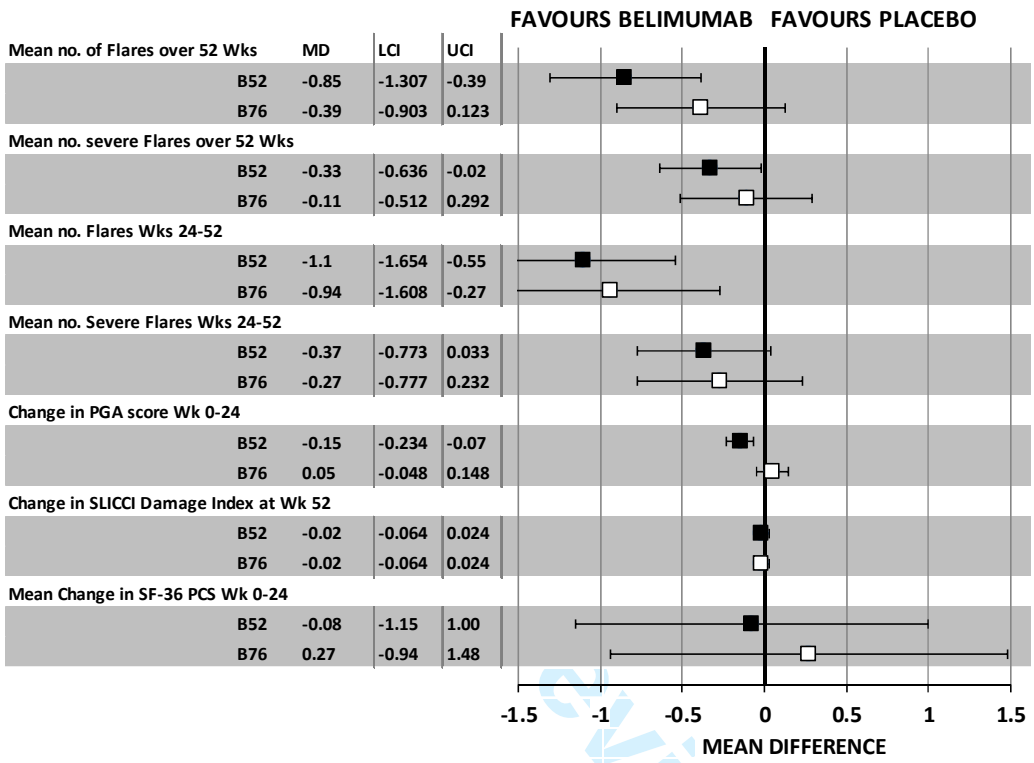


FIGURE 6: Meta-analysis of major outcomes in the two BLISS trials

Upper panel shows pooled estimates for binary and time to event outcomes (OR = odds ratio; HR = hazard ratio). Lower panel shows pooled estimates for continuous outcomes (MD = mean difference). SLICC = Systemic Lupus International Collaborating Clinics, the SLICC index is a measure of organ damage. Meta-analysis was conducted using random effects method (DerSimonian Laird).

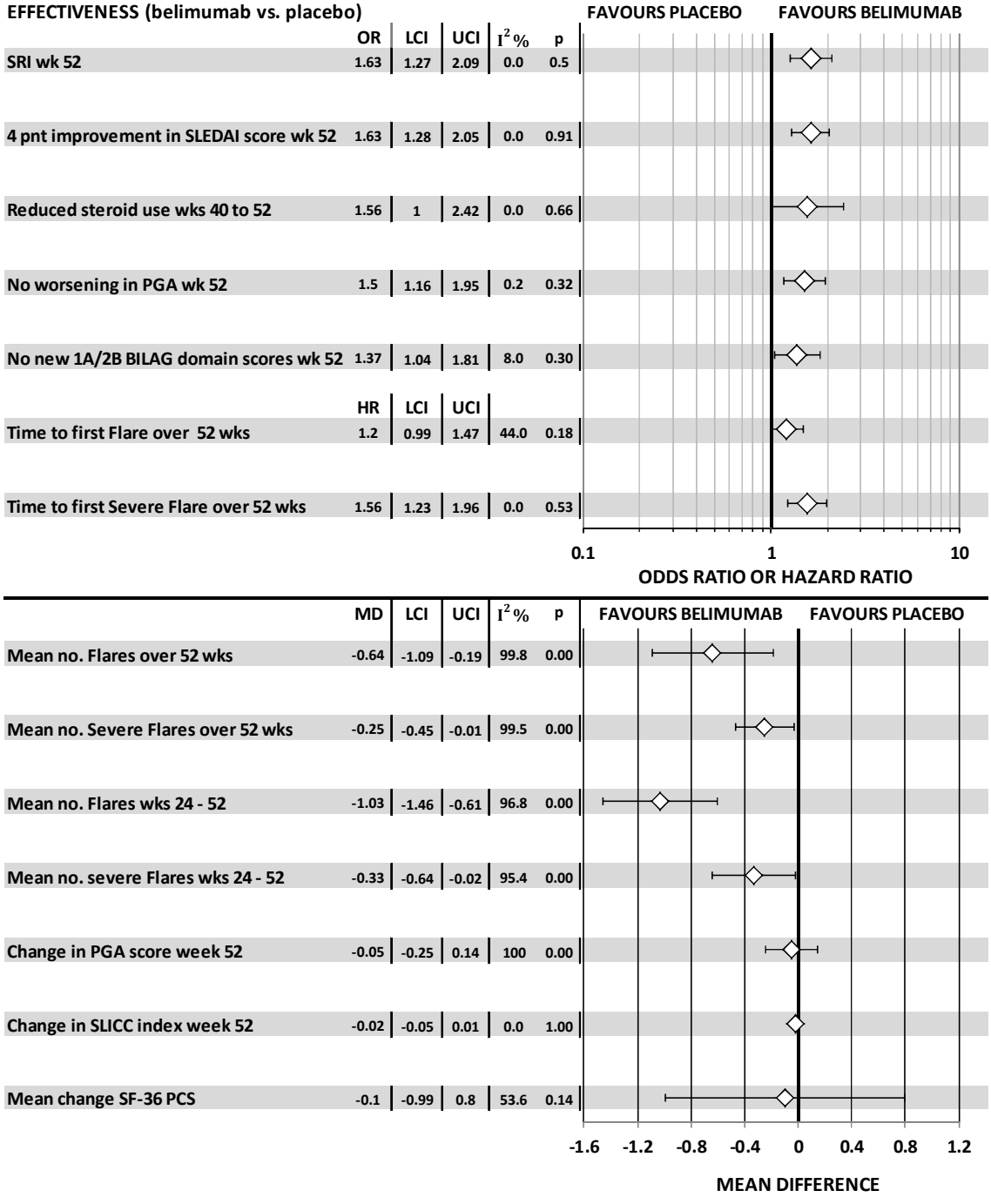


Table 1 Quality assessment of the included trials

QUALITY ITEMS	L02	BLISS-52	BLISS-76
Does reporting suggest that randomisation was carried out appropriately?	Yes	Yes	Yes
Does reporting suggest that the concealment of treatment allocation adequate?	Unclear	Unclear	Unclear
Were the groups reported as similar at the outset of the study in terms of prognostic factors?	Yes	Yes	Yes
Were the care providers, participants and outcome assessors reported as blind to treatment allocation?	Yes	Yes	Yes
Were any unexpected imbalances in drop-outs reported between groups?	No	No	No
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No	No	No
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	Yes	Yes

Quality assessment used information presented in the study journal articles and the manufacturer's submission to the US FDA and was based on CRD guidance (2008)[19] for undertaking systematic reviews in health care (CRD = Centre for Reviews and Dissemination. York: Centre for Reviews and Dissemination)

Table 2: Major characteristics of included studies

STUDY	Treatment (IV)	N	Mean Age (SD) yrs	SELENA-SLEDAI at entry	Geographical distribution of patients	Ethnic make-up of trial participants			Number and location of STUDY CENTRES
L02 2006 Phase II 52 week	Bel 1 mg/kg Bel 4 mg/kg Bel 10 mg/kg Placebo	114 111 111 113	42 (11)	> 4 points	US (98%), Canada (2%)	Caucasian	NR	69.9%	59 in N. America
						African American	NR	24.7%	
						Latino	NR	18.5%	
BLISS-52 2009 Phase III 52 week	Bel 1 mg/kg Bel 10 mg/kg Placebo	288 290 287	36 (11)	> 6 points	Latin America (50%), Asia (38%), E Europe & Australia (13%)	Caucasian	229	27%	90 in Pacific Asia. 11 in S. America & E. Europe
						Asian	327	38%	
						Black/African Am	30	4%	
						Alaskan Nat./Am Indian	279	32%	
						Nat. Hawaiian/Pacific Islander	0	0%	
						Multiracial	5	1%	
BLISS-76 2009 Phase III 76 week	Bel 1 mg/kg Bel 10 mg/kg Placebo	271 273 275	40 (12)	> 6 points	US & Canada (53%), W Europe (25%) E Europe (11%) Latin America (11%)	Caucasian	569	70%	136 in N. America & Europe
						Asian	28	3%	
						Black/African Am	118	14%	
						Alaskan Nat./Am Indian	103	13%	
						Nat. Hawaiian/Pacific Islander	1	0%	
						Multiracial	8	1%	

NR = not reported

Table 3: Outcomes defined and pre specified in the BLISS 52 and BLISS 76 trials and their accompanying designated status

Outcome	Measure	Outcome specification
SLE Responder Index (SRI*)	% responders at wk 52	Primary outcome
Reduction in SLEDAI score by ≥ 4 points	% responders at wk 52	Major secondary outcome
Change in PGA score from baseline	*Mean change at wk 24*	*Major secondary outcome*
Steroid reduction weeks 40 to 52	% responders	Major secondary outcome
SF-36 Physical component summary score	*Mean change at wk 24*	*Major secondary outcome*
SLE Responder Index	% responders at week 76	Major secondary outcome
SLICC/ACR damage index	*Mean change at wk 52*	*Secondary outcome*
FACIT-fatigue scale mean change from baseline	*Mean change at clinic visits*	*Secondary outcome*
EQ-5D score	*Mean change at clinic visits*	*Secondary outcome*
Change in PGA score from baseline	*Mean change at wk 52*	*Secondary outcome*
SF-36 Physical component summary score	*Mean change at wk 52*	*Secondary outcome*
SLEDAI SLE flare index over 52 wks	Time to first flare	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*
Death	Number during exposure	Safety outcome
Treatment emergent adverse events	Number during exposure	Safety outcome
Serious infections	Number during exposure	Safety outcome

* 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.

Continuous outcomes are in italics.

Appendix 1

Search Strategies

Cochrane Central Register of Controlled Trials (CENTRAL)

CENTRAL searched via Cochrane Library Interface on 18/05/11

1	MeSH descriptor Lupus Erythematosus, Systemic explode all trees	418
2	(lupus NEAR/3 erythematosus) or (systemic* NEAR/3 lupus) or (SLE)	630
3	(#1 OR #2)	703
4	belimumab OR benlysta	6
5	(#3 AND #4)	4

Medline

Medline searched via Ovid Interface on 19/05/11

1	exp Lupus Erythematosus, Systemic/	42025
2	(lupus adj3 erythematosus).tw.	35497
3	(systemic* adj3 lupus).tw.	31639
4	1 or 2 or 3	50358
5	belimumab.mp.	68
6	benlysta.mp.	3
7	5 or 6	68
8	4 and 7	48
9	randomized controlled trial.pt.	305892
10	controlled clinical trial.pt.	82328
11	randomized.ab.	212836
12	placebo.ab.	124063
13	clinical trials as topic.sh.	153987
14	randomly.ab.	154440
15	trial.ti.	91188
16	9 or 10 or 11 or 12 or 13 or 14 or 15	711420
17	exp animals/ not humans.sh.	3582822
18	16 not 17	656689
19	8 and 18	24

RCT search filter used: Cochrane Highly Sensitive Search Strategy for identifying randomized trials in MEDLINE: sensitivity- and precision-maximizing version (2008 revision); Ovid format. Box 6.4.b in the Cochrane handbook. Reference: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org

Medline In-process

Medline In-Process searched via Ovid Interface on 19/05/11

1	exp Lupus Erythematosus, Systemic/	0
2	(lupus adj3 erythematosus).tw.	1213
3	(systemic* adj3 lupus).tw.	873
4	1 or 2 or 3	1236
5	belimumab.mp.	8
6	benlysta.mp.	4
7	5 or 6	10
8	4 and 7	6

Embase

1	belimumab.mp.orexpbelimumab/	427
2	benlysta.mp.	24
3	1 or 2	428
4	exp systemic lupus erythematosus/	50906
5	(lupus adj3 erythematosus).tw.	40637
6	(systemic: adj3 lupus).tw.	36554
7	4 or 5 or 6	59739
8	3 and 7	302
9	random:.tw.	632763
10	placebo:.mp.	250140
11	double-blind:.tw.	116148
12	9 or 10 or 11	796900
13	8 and 12	144

RCT search filter used: Wong, et al. (2006) Best optimization of sensitivity and specificity. Reference: Wong SS, Wilczynski NL, Haynes RB. Developing optimal search strategies for detecting clinically sound treatment studies in EMBASE. J Med Libr Assoc. 2006 Jan;94(1):41-7. PubMed PMID: 16404468; PubMed Central PMCID: PMC1324770.

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1-2
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	3
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	3
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	3-4
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	4
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Appendix 1
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	4
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	4
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	4
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	4
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	5
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	5

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	4
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	NA
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	5
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	Table 2 and Figure 3
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	Table 1
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Figure 1 and Table 3
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	See Figure 6
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	Table 1
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	NA
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	11
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	8-11
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	10
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	11

PRISMA 2009 Checklist

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: **www.prisma-statement.org**.

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	Page 1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	Page 2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	Page 3-4
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	Page 3-4
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	Exists, available from authors
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	Page 5-6
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	Page 5-6
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Appendix
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	Page 5-6
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	Page 5-6
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	Page 5-6
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	Page 8-10
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	Page 8-12
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	Page 5-6

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	Page 5-6
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	NA
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	Fig 2
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	Page 8-12
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	Page 8-12
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Figures
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	Tables
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	Discussion
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	NA
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	Discussion
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	Discussion
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	Discussion
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	End of paper

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: **www.prisma-statement.org**.

Belimumab: a technological advance for SLE patients? Report of a systematic review and meta-analysis

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2013-002852.R2
Article Type:	Research
Date Submitted by the Author:	19-Jun-2013
Complete List of Authors:	Kandala, Ngianga-Bakwin; University of Warwick, Warwick Medical School; University of Oxford, KEMRI-University of Oxford-Wellcome Trust Collaborative Programme,Malaria Public Health and Epidemiology Group, Centre for Geographic Medicine Connock, Martin; University of Warwick, Division of Health Sciences, Warwick Medical School Grove, Amy; University of Warwick, Division of Health Sciences, Warwick Medical School Sutcliffe, Paul; University of Warwick, Division of Health Sciences, Warwick Medical School Mohiuddin, Syed; University of Warwick, Division of Health Sciences, Warwick Medical School Hartley, Louise; University of Warwick, Division of Health Sciences, Warwick Medical School Court, Rachel; Warwick University, Division of Health Sciences Cummis, Ewen; McMDCLtd, UK, G12 9TJ, McMaster Development Consultants Gordon, Caroline; University of Birmingham, School of Immunity and Infection, College of Medical and Dental Sciences Clarke, Aileen; University of Warwick, Division of Health Sciences
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Immunology (including allergy), Evidence based practice, Public health, Pharmacology and therapeutics
Keywords:	CLINICAL PHARMACOLOGY, EPIDEMIOLOGY, PREVENTIVE MEDICINE, PUBLIC HEALTH

Belimumab: a technological advance for Systemic Lupus Erythematosus **patients?**
Report of a systematic review and meta-analysis

Ngianga-Bakwin Kandala, Principal Research Fellow[1, 5]
Martin Connock, Senior Research Fellow[1]
Amy Grove, Project Manager[1]
Paul Sutcliffe, Associate Professor[1]
Syed Mohiuddin, Research Fellow[2]
Louise Hartley, Research Associate[1]
Rachel Court, Information Specialist[1]
Ewen Cummins, Health Economist[3]
Caroline Gordon, Professor of Rheumatology and Consultant Rheumatologist[4]
Aileen Clarke, Professor of Public Health & Health Services Research[1]

[1]Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry, UK, CV4 7AL
[2]The University of Manchester, Health Sciences, Manchester, UK, M13 9PL
[3]McMDCLtd, McMaster Development Consultants, Glasgow, UK, G12 9TJ
[4]School of Immunity and Infection, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK, B15 2TT
[5]KEMRI-University of Oxford-Wellcome Trust Collaborative Programme, Malaria Public Health and Epidemiology Group, Centre for Geographic Medicine, Nairobi, Kenya.

Correspondence to:
Ngianga-Bakwin Kandala, PhD
Principal Research Fellow in Health Technology Assessment
Warwick Evidence and Populations Evidence and Technologies
Division of Health Sciences
Warwick Medical School
University of Warwick
Coventry, CV4 7AL
Tel: +44 (0) 24761 50541
Email: N-B.Kandala@warwick.ac.uk

Short Title:

Systematic review on belimumab for SLE

ARTICLE SUMMARY

ARTICLE FOCUS

- SLE is a complex multi-organ auto-immune disease subject to relapse and remission.
- Patients almost always have fatigue, skin rashes and arthritis and there is a wide variety of other problems which the disease can cause.
- Belimumab is a new treatment specifically targeted against SLE.

KEY MESSAGES

1. Combining the results from two RCTs suggests that belimumab is clinically effective for SLE patients.
2. However, all outcomes were systematically superior in one trial compared with the other.
3. Different trial conduct and populations mean that it is too early to draw generalisable conclusions.

STRENGTHS AND LIMITATIONS

- At first sight combined meta analytic evidence suggests that belimumab is clinically effective for patients with severe SLE.
- We suggest that it is too early to draw strong conclusions because the two relevant trials cover different populations in different countries and there may be differences in trial conduct and outcome assessment.

Abstract:

Objectives: To undertake a systematic review and meta-analysis to investigate clinical effectiveness of belimumab for patients with SLE and anti-nuclear and/or anti-dsDNA autoantibodies.

Methods: We searched eight electronic databases and reference lists for randomised controlled trials (RCTs) of belimumab against placebo or best supportive care. Quality assessment and random effects meta-analysis were undertaken.
Design: A meta-Analysis of RCTs.
Setting: NA
Participants: 2133 SLE patients
Interventions: NA
Primary and secondary outcome measures: Responder Index (SRI) at week 52.

Results: Three double-blind placebo-controlled RCTs (L02, BLISS-52 BLISS-76) investigated 2133 SLE patients. BLISS-52 and BLISS-76 trials recruited patients with anti-nuclear and/or anti-dsDNA autoantibodies and demonstrated belimumab effectiveness for the SLE Responder Index (SRI) at week 52. Ethnicity and geographical location of participants varied considerably between BLISS trials. Although tests for statistical heterogeneity were negative, BLISS-52 results were systematically more favourable for all measured outcomes. Meta-analysis of pooled 52-week SRI BLISS results showed benefit for belimumab (OR 1.63, 95% CI 1.27-2.09). By week 76, the primary SRI outcome in BLISS-76 was not statistically significant (OR 1.31, 95% CI 0.919-1.855).

INTRODUCTION

Systemic Lupus Erythematosus (SLE) is an auto-immune disease subject to relapse and remission. Incidence is estimated at between 1.0 and 10.0 per hundred thousand person years using different measures, and prevalence at between 20-70 per 100,000.[1;2] SLE is a complex multi-organ disease with a number of different manifestations.[3] Patients almost always have fatigue, often have skin rashes and arthritis and there is a wide variety of other problems which the disease can cause.

The American College of Rheumatology has defined 11 classification criteria, including: rash; photosensitivity; oral ulcers; arthritis; serositis; renal and neurological disorder.[4;5] Assessment of the patient can be difficult, as flares of the disease have to be distinguished from its complications, from comorbidity especially infection, and from adverse effects of medications.[6] SLE is more common in women (in most studies 90% or more of cases are women[2]) and in those from black and other ethnic groups. Recently age-adjusted incidence rates have been produced showing that rates are highest in women aged 40 years and

over.[7] Mortality rates show that five year survival is high, at over 90%[8;9] and an overall SMR has been calculated as 2.4.[10]

Antinuclear antibodies are present in virtually all patients with SLE.[11] Anti-ds DNA antibodies are present in 50-60% patients at some point in their disease but often transiently with active disease.[11] Corticosteroids are the mainstay of treatment, they suppress disease but they may cause organ damage. The aim of treatment is to maintain normal function whilst suppressing disease activity and preventing organ damage,[6] achieving these conflicting aims can be difficult. Other drugs used include antimalarials such as hydroxychloroquine, and immunosuppressive drugs such as azathioprine and mycophenolatemofetil. More recently rituximab (a monoclonal antibody which reacts with the CD20 antigen expressed on B cells) has also been used, although the largest trial undertaken to date failed to reach its end point.[12]

Belimumab (Benlysta®) is an IgG1monoclonal antibody which inhibits the activity of the soluble cytokine BLyS (B lymphocyte stimulator; also known as BAFF).[13] In contrast to earlier SLE treatments, belimumab is targeted at the fundamental pathology of SLE and has been widely interpreted as representing a step change in treatment options.[13]

Belimumab was recently licensed in the USA and in Europe for treatment of autoantibody-positive SLE and is the first drug to be so licensed for several decades. The European indication is for severely affected SLE patients with active, autoantibody-positive disease and a high degree of disease activity exemplified by positive anti-ds DNA and low complement despite standard therapy.[13] Belimumab is administered by IV infusion recommended at 10 mg belimumab / kg on days 0, 14 and 28, and at 28 day intervals thereafter. A course of belimumab treatment for a 64 kg patient using the US list price of $1,477 (£926.37) for a 400 mg vial[14] would be $56,527 (£35,454) per year, and according to the US average whole sale price of $4.432 (£2780) / 400 mg vial)[15] would be $42,545 (£26,684) per year.

A number of clinical measures have been developed for tracking the progression of SLE[16] and for estimating the effects of treatment.[17] They include the Physician's Global Assessment (PGA), the SELENA-SLEDAI (Safety of Estrogen in Lupus National Assessment-Systemic Lupus Erythematosus Disease Activity Index), the BILAG Index (British Isles Lupus Assessment Group Index), and the SRI index (SLE Response Index). Their major features are summarised in Figure 1. Their complexity means that outside specialised centres they may not be widely used in routine clinical practice. The multiplicity

of SLE manifestations and of the systems developed to measure them has resulted in a proliferation of outcome measures that can be reported in trials of interventions for SLE. This in turn means that by chance at least some outcome measures will generate favourable results for an intervention; hence the US Federal Drug agency (FDA) in conjunction with belimumab-trialists developed the SRI aimed at guarding against the possibility that worsening in overall disease might be masked by apparent improvement in a more narrowly defined manifestation.

[Insert Figure 1 here]

Our objective was to synthesise findings from randomised controlled trials (RCTs) of belimumab for patients with SLE and anti-nuclear and /or anti-ds DNA autoantibodies, to make an overall assessment of the performance of this drug in relation to comparator treatments using the SRI and other outcomes (as listed in Figure 1) and to assess the findings of trials in the light of population samples and geographical factors.[18]

METHODS

The study was undertaken as part of work for the National Institute for Health Research, Health Technology Assessment programme (Grant funding reference 10/73/01. Further information is available from:www.hta.ac.uk/).

Search scope

We searched for RCTs investigating belimumab administered i.v. for patients with SLE and anti-nuclear and /or anti-ds DNA autoantibodies. Comparators considered were belimumab versus placebo and belimumab versus best supportive care. Outcomes included all disease-related or health-status-related measures. There was no publication year restriction, but the search was restricted to English language references only.

Search strategy

The following eight databases were searched: Cochrane Database of Systematic Reviews; the Cochrane Central Register of Controlled Trials (CENTRAL); DARE; EMBASE; HTA Database; Medline; Pre-Medline and Science Citation Index. Search strategies for these databases used a combination of terms related to the population and interventions listed above; the specific search strategies are provided in Appendix 1. In Medline and EMBASE the subject strategies were combined with search strategies designed to identify RCTs. (Appendix 1).

Unpublished studies were identified using: Clinical Trials, Current Controlled Trials, EU Clinical Trials Register, UK Clinical Research Network Study Portfolio, National Research Register, WHO Clinical Trials Search Portal, NHS Evidence, Conference Proceedings Citation Index -Science and Google.

In addition, specific websites were searched: Medicines and Healthcare products Regulatory Agency (MHRA), European Medicines Agency (EMEA), US Food and Drug Administration (FDA) and the following specific conference proceedings: American College of Rheumatology, British Society of Rheumatology and the European League Against Rheumatism (EULAR).

Inclusion criteria: Publications were included if they described results from RCTs of belimumab for SLE patients with positive autoantibodies. Two reviewers independently assessed retrieved publications for inclusion. There were no disagreements between reviewers.

Date extraction: Potentially relevant publications were obtained in full text and assessed by the same two reviewers. One reviewer extracted data for all specified primary and secondary outcome measures, for adverse events and deaths. A second reviewer checked extracted data.

Quality evaluation: Quality assessment and risk of bias was guided by the Centre for Reviews and Dissemination (CRD) checklist[19] based on all information in the included publications which specifies reporting of randomisation, concealment of allocation, group balance, blinding, drop-outs, outcome reporting bias, and whether intention to treat analysis was used.

Statistical analysis: Unadjusted odds ratios (ORs) and mean differences were calculated for binary and continuous outcomes respectively. Statistical heterogeneity was calculated using the I^2 statistic.[20;21] There were too few studies for an analysis of publication bias.[21] Although our thorough search found no further studies, we cannot completely rule out that any method for combining the two trials may result in an over-estimate or under-estimate of effect sizes due to publication bias. Adjusted outcome measures were tabulated where these were reported. A random effects meta-analysis[22] was undertaken using the DerSimonian Laird method in STATA version 11..[23] All graphs were prepared in Microsoft Excel 2010.

RESULTS

Characteristics of included studies

We identified three placebo controlled RCTs of belimumab versus standard care: the phase III trials termed BLISS-52[24] and BLISS-76[25] and a phase II trial (study L02).[26] The PRISMA flow chart shows the process of identification of publications (see Figure 2). We identified an on-going trial in Asia.[27] All three completed trials appeared to be of good quality; however details of allocation concealment were meagre (Table 1). In meta-analysis we included the two phase III trials (BLISS-52 and BLISS-76) since the population, trial design and primary outcome was different in the L02 trial.

[Insert Table 1 here]

[Insert Figure 2 here]

BLISS-52,[24] BLISS-76[25] and study L02[26] have been published in peer reviewed journals, however the fullest accounts in the public domain are in the FDA licensing approval documents[28;29] and the manufacturer's 2011 submission to the UK National Institute of Health and Clinical Excellence (NICE).[30] Each of these placebo-controlled randomised trials was designed with multiple randomised groups. In the L02 trial, patients received 1 or 4 or 10 mg/kg of belimumab or placebo, while in the BLISS trials the belimumab dose regimens were 1mg/kg or 10 mg/kg. Both US and European licensing is for the 10mg/kg dose regimen. In this article we focus on efficacy results for the 10mg/kg licensed regimen relative to placebo. We also consider the off licence 1 mg/kg and 4 mg/kg dose regimens for investigation of adverse events.

Centralised, stratified randomisation was used in all three trials and arms were generally well balanced. For the phase III trials, stratification was undertaken according to race, baseline proteinuria and disease activity score (SELENA SLEDAI); for the phase I study, disease activity only was used as a stratification factor. All three trials recruited predominantly female patients (~90%) and were described as double blind. The two BLISS studies were conducted according to similar protocols.

There were differences in geographical distribution of the study centres and in the resulting ethnic racial make-up of the study populations (Table 2 and Figure 3). Thus in BLISS-76,

70% were Caucasian, 13% Native American and 3% Asian, respectively, whereas in BLISS-52, 27% were Caucasian, 32% native American and 38% were Asian. Table 3 lists the major protocol pre-specified outcomes in the BLISS trials.

There were additional population differences between BLISS and L02 trials at recruitment. Reporting of results for patients with anti-nuclear and /or anti-ds DNA autoantibodies in L02 was only included for a post-hoc subgroup and primary outcomes measured in L02 were not comparable with those of the BLISS studies. For these reasons, L02 study results are included here only with regard to safety outcomes. For the BLISS trials a composite novel primary outcome measure was developed *a priori* from discussions between the FDA and the manufacturer and termed the SLE Response Index (SRI) (see Figure 1 and Table 3). The protocol pre-specified primary end point was the proportion of SRI responders at week 52. This is taken as the primary outcome in this systematic review.

[Insert Table 2 here]

[Insert Figure 3 here]

[Insert Table 3 here]

[Insert Figure 4here]

Efficacy results in the two BLISS trials for major binary effectiveness outcomes including the time to first SLE flare and to first severe flare are summarised in Figure 4. ORs have been calculated using the proportions of patients with and without events reported in the journal articles for these trials.[24;25] Safety outcomes shown in Figure 4 were calculated after combining the number of events across the three trials (L02, BLISS-52 and BLISS-76) and are taken from the FDA documents. The hazard ratios (HRs) for time to flares were poorly reported in journal articles and the data presented are taken from the manufacturer's submission to the FDA.[28;29] As shown in Figure 4 both trials satisfied this primary end point with a better result for BLISS-52. The difference in percentage responders in the belimumab group relative to placebo group was larger in BLISS-52 (14%), than in in BLISS-76 (9.4%).

For the other binary effectiveness outcomes, the BLISS-52 trial delivered results which were more favourable to belimumab than did BLISS-76, with the latter results failing to reach a conventional level of statistical significance except for the ≥ 4 point improvement in SLEDAI

score at week 52. The journal articles and manufacturer's submissions to the FDA and to NICE used a logistic regression model and reported ORs adjusted according to the stratification factors employed at randomisation. Adjusted ORs 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 = 0.0207). Again a superior response was found for the BLISS-52 trial. By week 76, the unadjusted OR for the SRI response in the BLISS-76 trial ceased to reach statistical significance (Figure 4); this also held for the reported OR adjusted by logistic regression (OR 1.31, 95% CI: 0.92 – 1.87, p = 0.1323).[29]

With regard to time to first flare or to first severe flare (each estimated over 52 weeks follow up) the responses reported in the FDA submission are again superior for BLISS-52. Each outcome failed to reach conventional statistical significance for BLISS-76. The FDA submission additionally reported more mature results estimated over 76 weeks of follow up for BLISS-76, and again these indicate lack of statistical significance for both outcomes (HR for first flare: 1.05, 95% CI: 0.88 – 1.27; HR for first severe flare 1.30, 95% CI: 0.92 – 1.85).

Figure 4 shows the results for major safety outcomes. Although there were more serious adverse events, more serious infections and more deaths associated with belimumab than with placebo, none of the ORs for these outcomes reached statistical significance. There were 14 deaths during the controlled phase of the three trials; three in the placebo group (n=675), and 11 in the belimumab groups (n=1458) with six in the 10mg/kg and five in the 1mg/kg groups, respectively (odds ratio 11.7; 95% CI 0.474 to 6.124). The causes of death were various: five were infection-related, three were strokes, three cardiovascular events, two suicides, one cancer, one from SLE-related complications, and two were of unknown cause.

Results for continuous outcomes are summarised in Figure 5. Mean changes from baseline reported in the BLISS journal articles and in the manufacturer's submissions to the FDA and NICE have been used to generate a mean difference statistic (sometimes termed "weighted mean difference"[31]). These revealed superiority of response in BLISS-52 relative to BLISS-76 for all reported outcomes, a pattern similar to that for binary outcomes. Mean changes from baseline for FACIT-fatigue scores and for EQ-5D utility scores (not pictured) did not reach statistical significance and again improvement seen in BLISS-52 for these was superior to that seen in BLISS-76.

In summary, BLISS-52 showed a systematic superiority over BLISS-76 in apparent benefit of belimumab across the full range of test responses (binary, time to event and continuous),

which may reflect geographical differences between the trials (Table 2 and Figure 3). The primary outcome in BLISS-76 was achieved (adjusted OR 1.52, 95% CI 1.07 to 2.15) but large geographical differences within BLISS-76 were striking: rates of 32% (46 out of 145), and 35% (47 out of 136), for placebo and belimumab respectively, were reported for patients from North America and Canada (a < 3% greater response for belimumab), whereas for BLISS-76 patients outside these regions a > 15% greater response for belimumab over placebo was reported, 71 of 137 (51.8%) for belimumab and 47 of 130 (36.1%) for placebo. In comparison, the corresponding rates for patients from Latin America in BLISS-52 were 49% placebo (71 out of 145), and 61% belimumab (85 out of 140).

[Insert Figure 5 here]

The manufacturer's submissions to the FDA and to NICE combined results from the two BLISS trials by pooling the patients and applying the logistic regression model described above; for the primary outcome (proportion of SRI responders at week 52), the difference between the belimumab and placebo groups was 11.8%.[28]

An alternate method of combining trials by meta-analysis of study level results from the two BLISS trials showed a statistically significant benefit of belimumab for most main outcomes including SRI, SELENA-SLEDAI, worsening in PGA, steroid use, BILAG score or, time to first severe flare, and mean number of flares and severe flares over 52 weeks and weeks 24 to 52 (Figure 6). Tests for statistical heterogeneity of ORs and HR outcomes were not significant. This Meta-analysis offers an alternative to the manufacturer logistic regression and it is justified for two trials of substantial size (N=577 and N=548), however, these results, and those from pooling individual patient data from the two trials prior to logistic regression, mask the systematic difference between trials across all outcomes.

[Insert Figure 6 here]

DISCUSSION

We undertook a systematic review of the clinical effectiveness of belimumab, a new treatment targeted at systemic lupus erythematosus (SLE) patients with anti-nuclear and /or anti-ds DNA autoantibodies. We performed an extensive search and systematic review of both completed and on-going trials using a number of databases and by checking reference lists. Data were extracted independently and studies were quality assessed. Random effects meta-analysis was undertaken.

We identified three RCTs (L02, BLISS-52, BLISS-76) reporting data on over 2000 patients. In contrast to the BLISS trials, L02 recruited patients who were not necessarily current carriers of anti-nuclear or anti ds DNA antibodies at study commencement. L02 failed to demonstrate clinical effectiveness for its primary end points.[26] Meta-analysis of the BLISS studies showed a benefit of belimumab with the main primary outcome (SRI), showing improvement at 52 weeks (OR 1.63; 95% CI: 1.27-2.09 p<0.001) although by week 76, the proportion of SRI responders in the BLISS-76 trial ceased to reach statistical significance (OR 1.31; 95% CI: 0.92–1.87 p=0.1323). There were no significant differences between placebo and intervention groups in quality of life or adverse events.

We found that the benefits of belimumab were systematically greater across the board (although not significantly so) in the BLISS-52 trial and although tests for statistical heterogeneity were negative, geographical location of study centres and the racial background and ethnicity of participants varied considerably. If the two BLISS trials were drawn from the same underlying populations, whilst one might expect outcomes to differ, we would anticipate that this would occur randomly between trials– some better some worse than the other.

A few studies have directly assessed the existence of and importance of geographical differences in trial outcomes.[32-34] Key factors contributing to such differences are variation in underlying patient population characteristics and variation in study execution. Vickers et al,[33] found that Eastern Asian and Eastern European studies had a higher proportion of positive trial results when compared to other countries. This is seen in the present case for the primary outcome where both the belimumab and placebo response rates in BLISS 52 study were greater than those in BLISS-76 and, remarkably, the placebo response rate in BLISS-52 (49%) was greater than that for the belimumab arm of BLISS-76 (43%). O'Shea and DeMets also report that within the Beta-Blocker Heart Attack Trial (BHAT), not only was there a difference in the direction, but also in the size of treatment effect between Canada and the US, although it should be noted that the original aim of that trial was not investigation of international differences in treatment effect.[35] One study found that 96-99% of the total variance in the "*Global utilisation of strategies to open occluded coronary arteries IV acute coronary syndromes*" (GUSTO IV ACS) trial could be accounted for by patient-level factors.[36]

International trials need to harmonise training of investigators, patient selection, treatment management, thresholds to centre admission, access to facilities, ascertainment of

endpoints and, by implication, results of interest [37-44] since it is possible that in centres in different countries these factors may differ systematically.[37] Equally, underlying differences in populations and countries (ethnicity, genetics, socio-economic status and health-care systems), and the nature and epidemiology of SLE according to ethnic background may result in differences in reporting of outcomes and pooled results.

The outcomes used in the BLISS trials would be unfamiliar to most of the investigators and it is possible that criteria may have differed between countries. In particular the Physician Global Assessment (PGA) is an important element of the outcomes measured (see Figure 1). PGA was measured as an outcome in itself, and it is also incorporated in SRI. PGA is of concern because as a global physician assessment of a patient's SLE status, it is subjective. The investigators reported a nearly 10% difference between the BLISS-52 and BLISS-76 studies in estimates of percentage change in PGA score in intervention groups at week 24 compared to baseline and this single result in one of the two trials is likely to have had an important influence on findings of the effectiveness of belimumab in SLE patients.

The latest results of belimumab in patients with SLE (phase II study design, uncontrolled extension study) reported that of 449 patients with active SLE (USA/Canada) 177 (39.4%) patients remained on treatment after 7 years of therapy (i.e. approximately 1746 cumulative patients-years) and that this subgroup exhibited durable sustained improvement in SLE disease activity (SRI and PGA).[30]

CONCLUSIONS

In conclusion, systematic review and random effects meta-analysis of two RCTs of belimumab for patients with autoantibody positive SLE demonstrated positive results in the main outcome at week 52. However, in view of the different populations studied at different locations in the BLISS trials and the consistently superior results from one trial compared to the other, we consider that population heterogeneity, geographical differences and variation in trial conduct and outcome assessment may have played a role in influencing outcomes. However the generalisability of results pooled meta-analytically or by logistic regression should be viewed with caution and we suggest that it is too early to draw strong conclusions in this case.

Acknowledgements

The authors would like to thank the National Institute for Health Research, Health Technology Assessment programme for funding this work.

Funding statement

This work was supported by the National Institute for Health Research, Health Technology Assessment programme [grant number 10/73/01].

Competing interest statement

No conflicts of interest.

Contributions:

N-BK: Conception and design. Data analysis and interpretation. Drafting the article. Critical revisions for important intellectual content. Approval of final article for submission.

MC: Conception and design. Data analysis and interpretation. Literature review. Interpretation of results. Drafting the article. Critical revisions for important intellectual content. Approval of final article for submission.

AG: Interpretation of results. Critical revisions for important intellectual content.

PS: Literature review. Interpretation of results. Critical revisions for important intellectual content.

SM: Data analysis and interpretation. Interpretation of results. Critical revisions for important intellectual content.

LH: Literature review. Interpretation of results. Critical revisions for important intellectual content.

RC: Literature review. Critical revisions for important intellectual content.

EC: Interpretation of results. Critical revisions for important intellectual content.

CG: Interpretation of results. Critical revisions for important intellectual content.

AC: Conception and design. Interpretation of results. Drafting the article. Critical revisions for important intellectual content. Approval of final article for submission.

All authors read and approved the final manuscript.

Data sharing

No additional data available.

Figure legends:

FIGURE 1: Summary of the major clinical measures used in SLE trials

FIGURE 2: PRISMA 2009 Flow Diagram for Belimumab in SLE RCTs and on-going trials

FIGURE 3: Differing centre locations in the BLISS 52 and BLISS 76 multicentre trials

FIGURE 4: Summary of results for major binary and time to event outcomes in belimumab RCTs

FIGURE 5: Summary of results for major continuous outcomes in BLISS 52 and BLISS 76 trials

FIGURE 6: Meta-analysis of major outcomes in the two BLISS trials

Reference List

(1) Bernatsky S, Joseph L, Pineau CA, et al. A population-based assessment of systemic lupus erythematosus incidence and prevalence--results and implications of using administrative data for epidemiological studies. *Rheumatology (Oxford)* 2007; 46(12):1814-1818.

(2) Pons-Estel GJ, Alarcon GS, Scofield L, et al. Understanding the epidemiology and progression of systemic lupus erythematosus. *Semin Arthritis Rheum* 2010; 39(4):257-268.

(3) Gordon C, Li CK, Isenberg DA. Systemic lupus erythematosus. *Medicine* 2010; 38(2):73-80.

(4) Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1997; 40(9):1725.

(5) Tan EM, Cohen AS, Fries JF, et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1982; 25(11):1271-1277.

(6) Griffiths B, Mosca M, Gordon C. Assessment of patients with systemic lupus erythematosus and the use of lupus disease activity indices. *Best Practice and Research in Clinical Rheumatology* 2005; 19(5 SPEC. ISS.):685-708.

(7) Somers EC, Thomas SL, Smeeth L, et al. Incidence of systemic lupus erythematosus in the United Kingdom, 1990-1999. *Arthritis Rheum* 2007; 57(4):612-618.

(8) Pons-Estel GJ, Alarcon GS, McGwin G, Jr., et al. Protective effect of hydroxychloroquine on renal damage in patients with lupus nephritis: LXV, data from a multiethnic US cohort. *Arthritis Rheum* 2009; 61(6):830-839.

(9) Pons-Estel GJ, Gonzalez LA, Zhang J, et al. Predictors of cardiovascular damage in patients with systemic lupus erythematosus: data from LUMINA (LXVIII), a multiethnic US cohort. *Rheumatology (Oxford)* 2009; 48(7):817-822.

(10) Bernatsky S, Boivin JF, Joseph L, et al. Mortality in systemic lupus erythematosus. *Arthritis and Rheumatism* 2006; 54(8):2550-2557.

(11) Kurien BT, Scofield RH. Autoantibody determination in the diagnosis of systemic lupus erythematosus. *Scand J Immunol* 2006; 64(3):227-235.

(12) Merrill JT, Neuwelt CM, Wallace DJ, 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; 62(1):222-233.

(13) Summary of product characteristics. 2012 Available from: URL: ema.europa.eu/Find medicine/Human medicines/European Public Assessment Reports Accessed January 2012

(14) American College of Rheumatology. Belimumab for Systemic Lupus Erythematosus. 2011 Available from: URL: http://www.rheumatology.org/publications/hotline/2011_03_15_belimumab.asp Accessed January 2012.

(15) New Drug Bulletin. Belimumab (Benlysta - Human Genome Sciences, Inc). 2012 Available from: URL:http://healthcare.utah.edu/pharmacy/bulletins/NDB_228.pdf Accessed January 2012.

(16) Isenberg DA, Rahman A, Allen E, et al. BILAG 2004. Development and initial validation of an updated version of the British Isles Lupus Assessment Group's disease activity index for patients with systemic lupus erythematosus. *Rheumatology* 2005; 44(7):902-906.

(17) Gordon C, Bertsias G, Ioannidis JPA, et al. EULAR points to consider for conducting clinical trials in Systemic lupus erythematosus. *Ann Rheum Dis* 2009; 68(4):470-476.

(18) Burgos PI, McGwin G, Jr., Pons-Estel GJ, et al. US patients of Hispanic and African ancestry develop lupus nephritis early in the disease course: data from LUMINA, a multiethnic US cohort (LUMINA LXXIV). *Ann Rheum Dis* 2011; 70(2):393-394.

(19) Khan KS, Ter Riet G, Glanville J, et al. Reviews of Research on Effectiveness. CRD's Guidance for Carrying Out or Commissioning Reviews. 2 nd Edition. (ISBN 1900640201) CRD Report No. 4. York: NHS Centre for Reviews and Dissemination (CRD), University of York. 2001 Available from: URL:http://www.york.ac.uk/inst/crd/report4.htm; Accessed January 2012.

(20) Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327(7414):557-560.

(21) Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. 2012 Available from: URL:www.cochrane-handbook.org Accessed January 2012.

(22) DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; 7(3):177-188.

(23) STATA Statistical analysis and software. 2011 Available from: URL:http://www.stata.com/ Accessed April 2012.

(24) Navarra SV, Guzman RM, Gallacher AE, et al. Efficacy and safety of belimumab in patients with active systemic lupus erythematosus: a randomised, placebo-controlled, phase 3 trial. *Lancet* 2011; 377(9767):721-731.

(25) Furie R, Petri M, Zamani O, et al. A phase III, randomized, placebo-controlled study of belimumab, a monoclonal antibody that inhibits B lymphocyte stimulator, in patients with systemic lupus erythematosus. *Arthritis Rheum* 2011; 63(12):3918-3930.

(26) Wallace DJ, Stohl W, Furie RA, 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(9):1168-1178.

(27) GlaxoSmithKline. GSK1550188 A 52 week study of belimumab versus placebo in the treatment of subjects with SLE located in North East Asia. 2011 Available from: URL:http://clinicaltrials.gov/ct2/show/NCT01345253 Accessed January 2012.

(28) 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. 2010 Available from: URL:http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/ArthritisDrugsAdvisoryCommittee/UCM233581.pdf Accessed January 2012.

(29) U.S.Food and Drugs Administration. FDA briefing package. 2010 Available from: URL:http://www.fda.gov/downloads/advisorycommittees/committeesmeetingmaterials/drugs/arthritisdrugsadvisorycommittee/ucm233579.pdf Accessed January 2012.

(30) GlaxoSmithKline. Single Technology Appraisal (STA): Belimumab for the treatment of active autoantibody-positive systemic lupus erythematosus. 2011 Available from: URL:http://www.guidance.nice.org.uk/TA/live/13307/56490/56490.pdf Accessed January 2012.

(31) Egger M, Smith GD, Altman DG. Systematic Reviews in Health Care, Meta-analysis in context. 2 ed. BMJ Books; 2001.

(32) Pan Z, Trikalinos TA, Kavvoura FK, Lau J, Ioannidis JP. Local literature bias in genetic epidemiology: an empirical evaluation of the Chinese literature. *PLoS Med* 2005; 2(12):e334.

(33) Vickers A, Goyal N, Harland R, et al. Do certain countries produce only positive results? A systematic review of controlled trials. *Control Clin Trials* 1998; 19(2):159-166.

(34) Zhang D, Freemantle N, Cheng KK. Are randomized trials conducted in China or India biased? A comparative empirical analysis. *J Clin Epidemiol* 2011; 64(1):90-95.

(35) O'Shea JC, DeMets DL. Statistical issues relating to international differences in clinical trials. *Am Heart J* 2001; 142(1):21-28.

(36) Chang WC, Midodzi WK, Westerhout CM, et al. Are international differences in the outcomes of acute coronary syndromes apparent or real? A multilevel analysis. *J Epidemiol Community Health* 2005; 59(5):427-433.

(37) White HD. International differences: selection, noise, or real? *Eur Heart J* 2000; 21(5):339-342.

(38) Fox KA, Goodman S, Bigonzi F, et al. Inter-regional differences and outcome in unstable angina; analysis of the international ESSENCE trial. Efficacy and Safety of Subcutaneous Enoxaparin in Non-Q-wave Coronary Events. *Eur Heart J* 2000; 21(17):1433-1439.

(39) Sinnaeve PR, Van de Werf FJ. Global patterns of health care for acute coronary syndromes. *Curr Opin Cardiol* 2004; 19(6):625-630.

(40) Faergeman O, Kjekshus J, Cook T, et al. Differences in the treatment of coronary heart disease between countries as revealed in the Scandinavian Simvastatin Survival Study (4S). *Eur Heart J* 1998; 19(10):1531-1537.

(41) Berger JS, Stebbins A, Granger CB, et al. Initial aspirin dose and outcome among ST-elevation myocardial infarction patients treated with fibrinolytic therapy. *Circulation* 2008; 117(2):192-199.

(42) Welsh RC, Chang W, Goldstein P, et al. Time to treatment and the impact of a physician on prehospital management of acute ST elevation myocardial infarction: insights from the ASSENT-3 PLUS trial. *Heart* 2005; 91(11):1400-1406.

(43) Yusuf S, Flather M, Pogue J, et al. Variations between countries in invasive cardiac procedures and outcomes in patients with suspected unstable angina or myocardial infarction without initial ST elevation. OASIS (Organisation to Assess Strategies for Ischaemic Syndromes) Registry Investigators. *Lancet* 1998; 352(9127):507-514.

(44) Bleyer AJ, Hylander B, Sudo H, et al. An international study of patient compliance with hemodialysis. *JAMA* 1999; 281(13):1211-1213.

Table 1 Quality assessment of the included trials

QUALITY ITEMS	L02	BLISS-52	BLISS-76
Does reporting suggest that randomisation was carried out appropriately?	Yes	Yes	Yes
Does reporting suggest that the concealment of treatment allocation adequate?	Unclear	Unclear	Unclear
Were the groups reported as similar at the outset of the study in terms of prognostic factors?	Yes	Yes	Yes
Were the care providers, participants and outcome assessors reported as blind to treatment allocation?	Yes	Yes	Yes
Were any unexpected imbalances in drop-outs reported between groups?	No	No	No
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No	No	No
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	Yes	Yes

Quality assessment used information presented in the study journal articles and the manufacturer's submission to the US FDA and was based on CRD guidance (2008)[19] for undertaking systematic reviews in health care (CRD = Centre for Reviews and Dissemination. York: Centre for Reviews and Dissemination)

Table 2: Major characteristics of included studies

STUDY	Treatment (IV)	N	Mean Age (SD) yrs	SELENA-SLEDAI at entry	Geographical distribution of patients	Ethnic make-up of trial participants			Number and location of STUDY CENTRES
L02 2006 Phase II 52 week	Bel 1 mg/kg Bel 4 mg/kg Bel 10 mg/kg Placebo	114 111 111 113	42 (11)	> 4 points	US (98%), Canada (2%)	Caucasian	NR	69.9%	59 in N. America
						African American	NR	24.7%	
						Latino	NR	18.5%	
BLISS-52 2009 Phase III 52 week	Bel 1 mg/kg Bel 10 mg/kg Placebo	288 290 287	36 (11)	> 6 points	Latin America (50%), Asia (38%), E Europe & Australia (13%)	Caucasian	229	27%	90 in Pacific Asia. 11 in S. America & E. Europe
						Asian	327	38%	
						Black/African Am	30	4%	
						Alaskan Nat./Am Indian	279	32%	
						Nat. Hawaiian/Pacific Islander	0	0%	
						Multiracial	5	1%	
BLISS-76 2009 Phase III 76 week	Bel 1 mg/kg Bel 10 mg/kg Placebo	271 273 275	40 (12)	> 6 points	US & Canada (53%), W Europe (25%) E Europe (11%) Latin America (11%)	Caucasian	569	70%	136 in N. America & Europe
						Asian	28	3%	
						Black/African Am	118	14%	
						Alaskan Nat./Am Indian	103	13%	
						Nat. Hawaiian/Pacific Islander	1	0%	
						Multiracial	8	1%	

NR = not reported

Table 3: Outcomes defined and pre specified in the BLISS 52 and BLISS 76 trials and their accompanying designated status

Outcome	Measure	Outcome specification
SLE Responder Index (SRI*)	% responders at wk 52	Primary outcome
Reduction in SLEDAI score by ≥ 4 points	% responders at wk 52	Major secondary outcome
Change in PGA score from baseline	*Mean change at wk 24*	*Major secondary outcome*
Steroid reduction weeks 40 to 52	% responders	Major secondary outcome
SF-36 Physical component summary score	*Mean change at wk 24*	*Major secondary outcome*
SLE Responder Index	% responders at week 76	Major secondary outcome
SLICC/ACR damage index	*Mean change at wk 52*	*Secondary outcome*
FACIT-fatigue scale mean change from baseline	*Mean change at clinic visits*	*Secondary outcome*
EQ-5D score	*Mean change at clinic visits*	*Secondary outcome*
Change in PGA score from baseline	*Mean change at wk 52*	*Secondary outcome*
SF-36 Physical component summary score	*Mean change at wk 52*	*Secondary outcome*
SLEDAI SLE flare index over 52 wks	Time to first flare	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*
Death	Number during exposure	Safety outcome
Treatment emergent adverse events	Number during exposure	Safety outcome
Serious infections	Number during exposure	Safety outcome

* 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.

Continuous outcomes are in italics.

Belimumab: a technological advance for Systemic Lupus Erythematosus **patients? Report of a systematic review and meta-analysis**

Ngianga-Bakwin Kandala, Principal Research Fellow[1, 5]
Martin Connock, Senior Research Fellow[1]
Amy Grove, Project Manager[1]
Paul Sutcliffe, Associate Professor[1]
Syed Mohiuddin, Research Fellow[2]
Louise Hartley, Research Associate[1]
Rachel Court, Information Specialist[1]
Ewen Cummins, Health Economist[3]
Caroline Gordon, Professor of Rheumatology and Consultant Rheumatologist[4]
Aileen Clarke, Professor of Public Health & Health Services Research[1]

[1]Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry, UK, CV4 7AL
[2]The University of Manchester, Health Sciences, Manchester, UK, M13 9PL
[3]McMDCLtd, McMaster Development Consultants, Glasgow, UK, G12 9TJ
[4]School of Immunity and Infection, College of Medical and Dental Sciences, University of Birmingham, Birmingham, UK, B15 2TT
[5]KEMRI-University of Oxford-Wellcome Trust Collaborative Programme, Malaria Public Health and Epidemiology Group, Centre for Geographic Medicine, Nairobi, Kenya.

Correspondence to:
Ngianga-Bakwin Kandala, PhD
Principal Research Fellow in Health Technology Assessment
Warwick Evidence and Populations Evidence and Technologies
Division of Health Sciences
Warwick Medical School
University of Warwick
Coventry, CV4 7AL
Tel: +44 (0) 24761 50541
Email: N-B.Kandala@warwick.ac.uk

Short Title:

Systematic review on belimumab for SLE

INTRODUCTION

Systemic Lupus Erythematosus (SLE) is an auto-immune disease subject to relapse and remission. Incidence is estimated at between 1.0 and 10.0 per hundred thousand person years using different measures, and prevalence at between 20-70 per 100,000.[1;2] SLE is a complex multi-organ disease with a number of different manifestations.[3] Patients almost always have fatigue, often have skin rashes and arthritis and there is a wide variety of other problems which the disease can cause.

The American College of Rheumatology has defined 11 classification criteria, including: rash; photosensitivity; oral ulcers; arthritis; serositis; renal and neurological disorder.[4;5] Assessment of the patient can be difficult, as flares of the disease have to be distinguished from its complications, from comorbidity especially infection, and from adverse effects of medications.[6] SLE is more common in women (in most studies 90% or more of cases are women[2]) and in those from black and other ethnic groups. Recently age-adjusted incidence rates have been produced showing that rates are highest in women aged 40 years and over.[7] Mortality rates show that five year survival is high, at over 90%[8;9] and an overall SMR has been calculated as 2.4.[10]

Antinuclear antibodies are present in virtually all patients with SLE.[11] Anti-ds DNA antibodies are present in 50-60% patients at some point in their disease but often transiently with active disease.[11] Corticosteroids are the mainstay of treatment, they suppress disease but they may cause organ damage. The aim of treatment is to maintain normal function whilst suppressing disease activity and preventing organ damage,[6] achieving these conflicting aims can be difficult. Other drugs used include antimalarials such as hydroxychloroquine, and immunosuppressive drugs such as azathioprine and mycophenolatemofetil. More recently rituximab (a monoclonal antibody which reacts with the CD20 antigen expressed on B cells) has also been used, although the largest trial undertaken to date failed to reach its end point.[12]

Belimumab (Benlysta®) is an IgG1monoclonal antibody which inhibits the activity of the soluble cytokine BLyS (B lymphocyte stimulator; also known as BAFF).[13] In contrast to earlier SLE treatments, belimumab is targeted at the fundamental pathology of SLE and has been widely interpreted as representing a step change in treatment options.[13]

Belimumab was recently licensed in the USA and in Europe for treatment of autoantibody-positive SLE and is the first drug to be so licensed for several decades. The European indication is for severely affected SLE patients with active, autoantibody-positive disease and a high degree of disease activity exemplified by positive anti-ds DNA and low complement despite standard therapy.[13] Belimumab is administered by IV infusion recommended at 10 mg belimumab / kg on days 0, 14 and 28, and at 28 day intervals thereafter. A course of belimumab treatment for a 64 kg patient using the US list price of $1,477 (£926.37) for a 400 mg vial[14] would be $56,527 (£35,454) per year, and according to the US average whole sale price of $4.432 (£2780) / 400 mg vial)[15] would be $42,545 (£26,684) per year.

A number of clinical measures have been developed for tracking the progression of SLE[16] and for estimating the effects of treatment.[17] They include the Physician's Global Assessment (PGA), the SELENA-SLEDAI (Safety of Estrogen in Lupus National Assessment-Systemic Lupus Erythematosus Disease Activity Index), the BILAG Index (British Isles Lupus Assessment Group Index), and the SRI index (SLE Response Index). Their major features are summarised in Figure 1. Their complexity means that outside specialised centres they may not be widely used in routine clinical practice. The multiplicity of SLE manifestations and of the systems developed to measure them has resulted in a proliferation of outcome measures that can be reported in trials of interventions for SLE. This in turn means that by chance at least some outcome measures will generate favourable results for an intervention; hence the US Federal Drug agency (FDA) in conjunction with belimumab-trialists developed the SRI aimed at guarding against the possibility that worsening in overall disease might be masked by apparent improvement in a more narrowly defined manifestation.

[Insert Figure 1 here]

Our objective was to synthesise findings from randomised controlled trials (RCTs) of belimumab for patients with SLE and anti-nuclear and /or anti-ds DNA autoantibodies, to make an overall assessment of the performance of this drug in relation to comparator treatments using the SRI and other outcomes (as listed in Figure 1) and to assess the findings of trials in the light of population samples and geographical factors.[18]

METHODS

The study was undertaken as part of work for the National Institute for Health Research, Health Technology Assessment programme (Grant funding reference 10/73/01. Further information is available from:www.hta.ac.uk/).

Search scope

We searched for RCTs investigating belimumab administered i.v. for patients with SLE and anti-nuclear and /or anti-ds DNA autoantibodies. Comparators considered were belimumab versus placebo and belimumab versus best supportive care. Outcomes included all disease-related or health-status-related measures. There was no publication year restriction, but the search was restricted to English language references only.

Search strategy

The following eight databases were searched: Cochrane Database of Systematic Reviews; the Cochrane Central Register of Controlled Trials (CENTRAL); DARE; EMBASE; HTA Database; Medline; Pre-Medline and Science Citation Index. Search strategies for these databases used a combination of terms related to the population and interventions listed above; the specific search strategies are provided in Appendix 1. In Medline and EMBASE the subject strategies were combined with search strategies designed to identify RCTs. (Appendix 1).

Unpublished studies were identified using: Clinical Trials, Current Controlled Trials, EU Clinical Trials Register, UK Clinical Research Network Study Portfolio, National Research Register, WHO Clinical Trials Search Portal, NHS Evidence, Conference Proceedings Citation Index -Science and Google.

In addition, specific websites were searched: Medicines and Healthcare products Regulatory Agency (MHRA), European Medicines Agency (EMEA), US Food and Drug Administration (FDA) and the following specific conference proceedings: American College of Rheumatology, British Society of Rheumatology and the European League Against Rheumatism (EULAR).

Inclusion criteria: Publications were included if they described results from RCTs of belimumab for SLE patients with positive autoantibodies. Two reviewers independently assessed retrieved publications for inclusion. There were no disagreements between reviewers.

Date extraction: Potentially relevant publications were obtained in full text and assessed by the same two reviewers. One reviewer extracted data for all specified primary and secondary outcome measures, for adverse events and deaths. A second reviewer checked extracted data.

Quality evaluation: Quality assessment and risk of bias was guided by the Centre for Reviews and Dissemination (CRD) checklist[19] based on all information in the included publications which specifies reporting of randomisation, concealment of allocation, group balance, blinding, drop-outs, outcome reporting bias, and whether intention to treat analysis was used.

Statistical analysis: Unadjusted odds ratios (ORs) and mean differences were calculated for binary and continuous outcomes respectively. Statistical heterogeneity was calculated using the I^2 statistic.[20;21] There were too few studies for an analysis of publication bias.[21] Although our thorough search found no further studies, we cannot completely rule out that any method for combining the two trials may result in an over-estimate or under-estimate of effect sizes due to publication bias. Adjusted outcome measures were tabulated where these were reported. A random effects meta-analysis[22] was undertaken using the DerSimonian Laird method in STATA version 11..[23] All graphs were prepared in Microsoft Excel 2010.

RESULTS

Characteristics of included studies

We identified three placebo controlled RCTs of belimumab versus standard care: the phase III trials termed BLISS-52[24] and BLISS-76[25] and a phase II trial (study L02).[26] The PRISMA flow chart shows the process of identification of publications (see Figure 2). We identified an on-going trial in Asia.[27] All three completed trials appeared to be of good quality; however details of allocation concealment were meagre (Table 1). In meta-analysis we included the two phase III trials (BLISS-52 and BLISS-76) since the population, trial design and primary outcome was different in the L02 trial.

[Insert Table 1 here]

[Insert Figure 2 here]

BLISS-52,[24] BLISS-76[25] and study L02[26] have been published in peer reviewed journals, however the fullest accounts in the public domain are in the FDA licensing approval documents[28;29] and the manufacturer's 2011 submission to the UK National Institute of Health and Clinical Excellence (NICE).[30] Each of these placebo-controlled randomised trials was designed with multiple randomised groups. In the L02 trial, patients received 1 or 4 or 10 mg/kg of belimumab or placebo, while in the BLISS trials the belimumab dose regimens were 1mg/kg or 10 mg/kg. Both US and European licensing is for the 10mg/kg dose regimen. In this article we focus on efficacy results for the 10mg/kg licensed regimen relative to placebo. We also consider the off licence 1 mg/kg and 4 mg/kg dose regimens for investigation of adverse events.

Centralised, stratified randomisation was used in all three trials and arms were generally well balanced. For the phase III trials, stratification was undertaken according to race, baseline proteinuria and disease activity score (SELENA SLEDAI); for the phase I study, disease activity only was used as a stratification factor. All three trials recruited predominantly female patients (~90%) and were described as double blind. The two BLISS studies were conducted according to similar protocols.

There were differences in geographical distribution of the study centres and in the resulting ethnic racial make-up of the study populations (Table 2 and Figure 3). Thus in BLISS-76, 70% were Caucasian, 13% Native American and 3% Asian, respectively, whereas in BLISS-52, 27% were Caucasian, 32% native American and 38% were Asian. Table 3 lists the major protocol pre-specified outcomes in the BLISS trials.

There were additional population differences between BLISS and L02 trials at recruitment. Reporting of results for patients with anti-nuclear and /or anti-ds DNA autoantibodies in L02 was only included for a post-hoc subgroup and primary outcomes measured in L02 were not comparable with those of the BLISS studies. For these reasons, L02 study results are included here only with regard to safety outcomes. For the BLISS trials a composite novel primary outcome measure was developed *a priori* from discussions between the FDA and the manufacturer and termed the SLE Response Index (SRI) (see Figure 1 and Table 3). The protocol pre-specified primary end point was the proportion of SRI responders at week 52. This is taken as the primary outcome in this systematic review.

[Insert Table 2 here]

[Insert Figure 3 here]

[Insert Table 3 here]

[Insert Figure 4here]

Efficacy results in the two BLISS trials for major binary effectiveness outcomes including the time to first SLE flare and to first severe flare are summarised in Figure 4. ORs have been calculated using the proportions of patients with and without events reported in the journal articles for these trials.[24;25] Safety outcomes shown in Figure 4 were calculated after combining the number of events across the three trials (L02, BLISS-52 and BLISS-76) and are taken from the FDA documents. The hazard ratios (HRs) for time to flares were poorly reported in journal articles and the data presented are taken from the manufacturer's submission to the FDA.[28;29] As shown in Figure 4 both trials satisfied this primary end point with a better result for BLISS-52. The difference in percentage responders in the belimumab group relative to placebo group was larger in BLISS-52 (14%), than in in BLISS-76 (9.4%).

For the other binary effectiveness outcomes, the BLISS-52 trial delivered results which were more favourable to belimumab than did BLISS-76, with the latter results failing to reach a conventional level of statistical significance except for the ≥ 4 point improvement in SLEDAI score at week 52. The journal articles and manufacturer's submissions to the FDA and to NICE used a logistic regression model and reported ORs adjusted according to the stratification factors employed at randomisation. Adjusted ORs 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 = 0.0207). Again a superior response was found for the BLISS-52 trial. By week 76, the unadjusted OR for the SRI response in the BLISS-76 trial ceased to reach statistical significance (Figure 4); this also held for the reported OR adjusted by logistic regression (OR 1.31, 95% CI: 0.92 – 1.87, p = 0.1323).[29]

With regard to time to first flare or to first severe flare (each estimated over 52 weeks follow up) the responses reported in the FDA submission are again superior for BLISS-52. Each outcome failed to reach conventional statistical significance for BLISS-76. The FDA submission additionally reported more mature results estimated over 76 weeks of follow up for BLISS-76, and again these indicate lack of statistical significance for both outcomes (HR for first flare: 1.05, 95% CI: 0.88 – 1.27; HR for first severe flare 1.30, 95% CI: 0.92 – 1.85).

Figure 4 shows the results for major safety outcomes. Although there were more serious adverse events, more serious infections and more deaths associated with belimumab than with placebo, none of the ORs for these outcomes reached statistical significance. There were 14 deaths during the controlled phase of the three trials; three in the placebo group (n=675), and 11 in the belimumab groups (n=1458) with six in the 10mg/kg and five in the 1mg/kg groups, respectively (odds ratio 11.7; 95% CI 0.474 to 6.124). The causes of death were various: five were infection-related, three were strokes, three cardiovascular events, two suicides, one cancer, one from SLE-related complications, and two were of unknown cause.

Results for continuous outcomes are summarised in Figure 5. Mean changes from baseline reported in the BLISS journal articles and in the manufacturer's submissions to the FDA and NICE have been used to generate a mean difference statistic (sometimes termed "weighted mean difference"[31]). These revealed superiority of response in BLISS-52 relative to BLISS-76 for all reported outcomes, a pattern similar to that for binary outcomes. Mean changes from baseline for FACIT-fatigue scores and for EQ-5D utility scores (not pictured) did not reach statistical significance and again improvement seen in BLISS-52 for these was superior to that seen in BLISS-76.

In summary, BLISS-52 showed a systematic superiority over BLISS-76 in apparent benefit of belimumab across the full range of test responses (binary, time to event and continuous), which may reflect geographical differences between the trials (Table 2 and Figure 3). The primary outcome in BLISS-76 was achieved (adjusted OR 1.52, 95% CI 1.07 to 2.15) but large geographical differences within BLISS-76 were striking: rates of 32% (46 out of 145), and 35% (47 out of 136), for placebo and belimumab respectively, were reported for patients from North America and Canada (a < 3% greater response for belimumab), whereas for BLISS-76 patients outside these regions a > 15% greater response for belimumab over placebo was reported, 71 of 137 (51.8%) for belimumab and 47 of 130 (36.1%) for placebo. In comparison, the corresponding rates for patients from Latin America in BLISS-52 were 49% placebo (71 out of 145), and 61% belimumab (85 out of 140).

[Insert Figure 5 here]

The manufacturer's submissions to the FDA and to NICE combined results from the two BLISS trials by pooling the patients and applying the logistic regression model described above; for the primary outcome (proportion of SRI responders at week 52), the difference between the belimumab and placebo groups was 11.8%.[28]

An alternate method of combining trials by meta-analysis of study level results from the two BLISS trials showed a statistically significant benefit of belimumab for most main outcomes including SRI, SELENA-SLEDAI, worsening in PGA, steroid use, BILAG score or, time to first severe flare, and mean number of flares and severe flares over 52 weeks and weeks 24 to 52 (Figure 6). Tests for statistical heterogeneity of ORs and HR outcomes were not significant. This Meta-analysis offers an alternative to the manufacturer logistic regression and it is justified for two trials of substantial size (N=577 and N=548), however, these results, and those from pooling individual patient data from the two trials prior to logistic regression, mask the systematic difference between trials across all outcomes.

[Insert Figure 6 here]

DISCUSSION

We undertook a systematic review of the clinical effectiveness of belimumab, a new treatment targeted at systemic lupus erythematosus (SLE) patients with anti-nuclear and /or anti-ds DNA autoantibodies. We performed an extensive search and systematic review of both completed and on-going trials using a number of databases and by checking reference lists. Data were extracted independently and studies were quality assessed. Random effects meta-analysis was undertaken.

We identified three RCTs (L02, BLISS-52, BLISS-76) reporting data on over 2000 patients. In contrast to the BLISS trials, L02 recruited patients who were not necessarily current carriers of anti-nuclear or anti ds DNA antibodies at study commencement. L02 failed to demonstrate clinical effectiveness for its primary end points.[26] Meta-analysis of the BLISS studies showed a benefit of belimumab with the main primary outcome (SRI), showing improvement at 52 weeks (OR 1.63; 95% CI: 1.27-2.09 p<0.001) although by week 76, the proportion of SRI responders in the BLISS-76 trial ceased to reach statistical significance (OR 1.31; 95% CI: 0.92–1.87 p=0.1323). There were no significant differences between placebo and intervention groups in quality of life or adverse events.

We found that the benefits of belimumab were systematically greater across the board (although not significantly so) in the BLISS-52 trial and although tests for statistical heterogeneity were negative, geographical location of study centres and the racial background and ethnicity of participants varied considerably. If the two BLISS trials were drawn from the same underlying populations, whilst one might expect outcomes to differ, we

would anticipate that this would occur randomly between trials– some better some worse than the other.

A few studies have directly assessed the existence of and importance of geographical differences in trial outcomes.[32-34] Key factors contributing to such differences are variation in underlying patient population characteristics and variation in study execution. Vickers et al,[33] found that Eastern Asian and Eastern European studies had a higher proportion of positive trial results when compared to other countries. This is seen in the present case for the primary outcome where both the belimumab and placebo response rates in BLISS 52 study were greater than those in BLISS-76 and, remarkably, the placebo response rate in BLISS-52 (49%) was greater than that for the belimumab arm of BLISS-76 (43%). O'Shea and DeMets also report that within the Beta-Blocker Heart Attack Trial (BHAT), not only was there a difference in the direction, but also in the size of treatment effect between Canada and the US, although it should be noted that the original aim of that trial was not investigation of international differences in treatment effect.[35] One study found that 96-99% of the total variance in the "*Global utilisation of strategies to open occluded coronary arteries IV acute coronary syndromes*" (GUSTO IV ACS) trial could be accounted for by patient-level factors.[36]

International trials need to harmonise training of investigators, patient selection, treatment management, thresholds to centre admission, access to facilities, ascertainment of endpoints and, by implication, results of interest [37-44] since it is possible that in centres in different countries these factors may differ systematically.[37] Equally, underlying differences in populations and countries (ethnicity, genetics, socio-economic status and health-care systems), and the nature and epidemiology of SLE according to ethnic background may result in differences in reporting of outcomes and pooled results.

The outcomes used in the BLISS trials would be unfamiliar to most of the investigators and it is possible that criteria may have differed between countries. In particular the Physician Global Assessment (PGA) is an important element of the outcomes measured (see Figure 1). PGA was measured as an outcome in itself, and it is also incorporated in SRI. PGA is of concern because as a global physician assessment of a patient's SLE status, it is subjective. The investigators reported a nearly 10% difference between the BLISS-52 and BLISS-76 studies in estimates of percentage change in PGA score in intervention groups at week 24 compared to baseline and this single result in one of the two trials is likely to have had an important influence on findings of the effectiveness of belimumab in SLE patients.

The latest results of belimumab in patients with SLE (phase II study design, uncontrolled extension study) reported that of 449 patients with active SLE (USA/Canada) 177 (39.4%) patients remained on treatment after 7 years of therapy (i.e. approximately 1746 cumulative patients-years) and that this subgroup exhibited durable sustained improvement in SLE disease activity (SRI and PGA).[30]

CONCLUSIONS

In conclusion, systematic review and random effects meta-analysis of two RCTs of belimumab for patients with autoantibody positive SLE demonstrated positive results in the main outcome at week 52. However, in view of the different populations studied at different locations in the BLISS trials and the consistently superior results from one trial compared to the other, we consider that population heterogeneity, geographical differences and variation in trial conduct and outcome assessment may have played a role in influencing outcomes. However the generalisability of results pooled meta-analytically or by logistic regression should be viewed with caution and we suggest that it is too early to draw strong conclusions in this case.

ARTICLE FOCUS

- SLE is a complex multi-organ auto-immune disease subject to relapse and remission.
- Patients almost always have fatigue, skin rashes and arthritis and there is a wide variety of other problems which the disease can cause.
- Belimumab is a new treatment specifically targeted against SLE.

KEY MESSAGES

1. Combining the results from two RCTs suggests that belimumab is clinically effective for SLE patients.
2. However, all outcomes were systematically superior in one trial compared with the other.
3. Different trial conduct and populations mean that it is too early to draw generalisable conclusions.

STRENGTHS AND LIMITATIONS

- At first sight combined meta analytic evidence suggests that belimumab is clinically effective for patients with severe SLE.
- We suggest that it is too early to draw strong conclusions because the two relevant trials cover different populations in different countries and there may be differences in trial conduct and outcome assessment.

Acknowledgements

The authors would like to thank the National Institute for Health Research, Health Technology Assessment programme for funding this work.

Funding statement

This work was supported by the National Institute for Health Research, Health Technology Assessment programme [grant number 10/73/01].

Competing interest statement

No conflicts of interest.

Contributions:

N-BK: Conception and design. Data analysis and interpretation. Drafting the article. Critical revisions for important intellectual content. Approval of final article for submission.

MC: Conception and design. Data analysis and interpretation. Literature review. Interpretation of results. Drafting the article. Critical revisions for important intellectual content. Approval of final article for submission.

AG: Interpretation of results. Critical revisions for important intellectual content.

PS: Literature review. Interpretation of results. Critical revisions for important intellectual content.

SM: Data analysis and interpretation. Interpretation of results. Critical revisions for important intellectual content.

LH: Literature review. Interpretation of results. Critical revisions for important intellectual content.

RC: Literature review. Critical revisions for important intellectual content.

EC: Interpretation of results. Critical revisions for important intellectual content.

CG: Interpretation of results. Critical revisions for important intellectual content.

AC: Conception and design. Interpretation of results. Drafting the article. Critical revisions for important intellectual content. Approval of final article for submission.

All authors read and approved the final manuscript.

431

Reference List

(1) Bernatsky S, Joseph L, Pineau CA, Tamblyn R, Feldman DE, Clarke AE. A population-based assessment of systemic lupus erythematosus incidence and prevalence--results and implications of using administrative data for epidemiological studies. *Rheumatology (Oxford)* 2007; 46(12):1814-1818.

(2) Pons-Estel GJ, Alarcon GS, Scofield L, Reinlib L, Cooper GS. Understanding the epidemiology and progression of systemic lupus erythematosus. *Semin Arthritis Rheum* 2010; 39(4):257-268.

(3) Gordon C, Li CK, Isenberg DA. Systemic lupus erythematosus. *Medicine* 2010; 38(2):73-80.

(4) Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1997; 40(9):1725.

(5) Tan EM, Cohen AS, Fries JF, Masi AT, McShane DJ, Rothfield NF et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1982; 25(11):1271-1277.

(6) Griffiths B, Mosca M, Gordon C. Assessment of patients with systemic lupus erythematosus and the use of lupus disease activity indices. *Best Practice and Research in Clinical Rheumatology* 2005; 19(5 SPEC. ISS.):685-708.

(7) Somers EC, Thomas SL, Smeeth L, Schoonen WM, Hall AJ. Incidence of systemic lupus erythematosus in the United Kingdom, 1990-1999. *Arthritis Rheum* 2007; 57(4):612-618.

(8) Pons-Estel GJ, Alarcon GS, McGwin G, Jr., Danila MI, Zhang J, Bastian HM et al. Protective effect of hydroxychloroquine on renal damage in patients with lupus nephritis: LXV, data from a multiethnic US cohort. *Arthritis Rheum* 2009; 61(6):830-839.

(9) Pons-Estel GJ, Gonzalez LA, Zhang J, Burgos PI, Reveille JD, Vila LM et al. Predictors of cardiovascular damage in patients with systemic lupus erythematosus: data from LUMINA (LXVIII), a multiethnic US cohort. *Rheumatology (Oxford)* 2009; 48(7):817-822.

(10) Bernatsky S, Boivin JF, Joseph L, Manzi S, Ginzler E, Gladman DD et al. Mortality in systemic lupus erythematosus. *Arthritis and Rheumatism* 2006; 54(8):2550-2557.

(11) Kurien BT, Scofield RH. Autoantibody determination in the diagnosis of systemic lupus erythematosus. *Scand J Immunol* 2006; 64(3):227-235.

(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; 62(1):222-233.

(13) Summary of product characteristics. 2012 Available from: URL: ema.europa.eu/Find medicine/Human medicines/European Public Assessment Reports Accessed January 2012

(14) American College of Rheumatology. Belimumab for Systemic Lupus Erythematosus. 2011 Available from: URL: http://www.rheumatology.org/publications/hotline/2011_03_15_belimumab.asp Accessed January 2012.

(15) New Drug Bulletin. Belimumab (Benlysta - Human Genome Sciences, Inc). 2012 Available from: URL:http://healthcare.utah.edu/pharmacy/bulletins/NDB_228.pdf Accessed January 2012.

(16) Isenberg DA, Rahman A, Allen E, Farewell V, Akil M, Bruce IN et al. BILAG 2004. Development and initial validation of an updated version of the British Isles Lupus Assessment Group's disease activity index for patients with systemic lupus erythematosus. *Rheumatology* 2005; 44(7):902-906.

(17) Gordon C, Bertsias G, Ioannidis JPA, Boletis J, Bombardieri S, Cervera R et al. EULAR points to consider for conducting clinical trials in Systemic lupus erythematosus. *Ann Rheum Dis* 2009; 68(4):470-476.

(18) Burgos PI, McGwin G, Jr., Pons-Estel GJ, Reveille JD, Alarcon GS, Vila LM. US patients of Hispanic and African ancestry develop lupus nephritis early in the disease course: data from LUMINA, a multiethnic US cohort (LUMINA LXXIV). *Ann Rheum Dis* 2011; 70(2):393-394.

(19) Khan KS, Ter Riet G, Glanville J, Snowdon AJ, Kleijnen J. Reviews of Research on Effectiveness. CRD's Guidance for Carrying Out or Commissioning Reviews. 2 nd Edition. (ISBN 1900640201) CRD Report No. 4. York: NHS Centre for Reviews and Dissemination (CRD), University of York. 2001 Available from: URL:http://www.york.ac.uk/inst/crd/report4.htm; Accessed January 2012.

(20) Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327(7414):557-560.

(21) Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. 2012 Available from: URL:www.cochrane-handbook.org Accessed January 2012.

(22) DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; 7(3):177-188.

(23) STATA Statistical analysis and software. 2011 Available from: URL:http://www.stata.com/ Accessed April 2012.

(24) 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(9767):721-731.

(25) Furie R, Petri M, Zamani O, Cervera R, Wallace DJ, Tegzova D et al. A phase III, randomized, placebo-controlled study of belimumab, a monoclonal antibody that

inhibits B lymphocyte stimulator, in patients with systemic lupus erythematosus. *Arthritis Rheum* 2011; 63(12):3918-3930.

(26) 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(9):1168-1178.

(27) GlaxoSmithKline. GSK1550188 A 52 week study of belimumab versus placebo in the treatment of subjects with SLE located in North East Asia. 2011 Available from: URL:http://clinicaltrials.gov/ct2/show/NCT01345253 Accessed January 2012.

(28) 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. 2010 Available from: URL:http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/ArthritisDrugsAdvisoryCommittee/UCM233581.pdf Accessed January 2012.

(29) U.S.Food and Drugs Administration. FDA briefing package. 2010 Available from: URL:http://www.fda.gov/downloads/advisorycommittees/committeesmeetingmaterials/drugs/arthritisdrugsadvisorycommittee/ucm233579.pdf Accessed January 2012.

(30) GlaxoSmithKline. Single Technology Appraisal (STA): Belimumab for the treatment of active autoantibody-positive systemic lupus erythematosus. 2011 Available from: URL:http://www.guidance.nice.org.uk/TA/live/13307/56490/56490.pdf Accessed January 2012.

(31) Egger M, Smith GD, Altman DG. Systematic Reviews in Health Care, Meta-analysis in context. 2 ed. BMJ Books; 2001.

(32) Pan Z, Trikalinos TA, Kavvoura FK, Lau J, Ioannidis JP. Local literature bias in genetic epidemiology: an empirical evaluation of the Chinese literature. *PLoS Med* 2005; 2(12):e334.

(33) Vickers A, Goyal N, Harland R, Rees R. Do certain countries produce only positive results? A systematic review of controlled trials. *Control Clin Trials* 1998; 19(2):159-166.

(34) Zhang D, Freemantle N, Cheng KK. Are randomized trials conducted in China or India biased? A comparative empirical analysis. *J Clin Epidemiol* 2011; 64(1):90-95.

(35) O'Shea JC, DeMets DL. Statistical issues relating to international differences in clinical trials. *Am Heart J* 2001; 142(1):21-28.

(36) Chang WC, Midodzi WK, Westerhout CM, Boersma E, Cooper J, Barnathan ES et al. Are international differences in the outcomes of acute coronary syndromes apparent or real? A multilevel analysis. *J Epidemiol Community Health* 2005; 59(5):427-433.

(37) White HD. International differences: selection, noise, or real? *Eur Heart J* 2000; 21(5):339-342.

(38) Fox KA, Goodman S, Bigonzi F, Le L, V, Cohen M. Inter-regional differences and outcome in unstable angina; analysis of the international ESSENCE trial. Efficacy

and Safety of Subcutaneous Enoxaparin in Non-Q-wave Coronary Events. *Eur Heart J* 2000; 21(17):1433-1439.

(39) Sinnaeve PR, Van de Werf FJ. Global patterns of health care for acute coronary syndromes. *Curr Opin Cardiol* 2004; 19(6):625-630.

(40) Faergeman O, Kjekshus J, Cook T, Pyorala K, Wilhelmsen L, Thorgeirsson G et al. Differences in the treatment of coronary heart disease between countries as revealed in the Scandinavian Simvastatin Survival Study (4S). *Eur Heart J* 1998; 19(10):1531-1537.

(41) Berger JS, Stebbins A, Granger CB, Ohman EM, Armstrong PW, Van de Werf F et al. Initial aspirin dose and outcome among ST-elevation myocardial infarction patients treated with fibrinolytic therapy. *Circulation* 2008; 117(2):192-199.

(42) Welsh RC, Chang W, Goldstein P, Adgey J, Granger CB, Verheugt FW et al. Time to treatment and the impact of a physician on prehospital management of acute ST elevation myocardial infarction: insights from the ASSENT-3 PLUS trial. *Heart* 2005; 91(11):1400-1406.

(43) Yusuf S, Flather M, Pogue J, Hunt D, Varigos J, Piegas L et al. Variations between countries in invasive cardiac procedures and outcomes in patients with suspected unstable angina or myocardial infarction without initial ST elevation. OASIS (Organisation to Assess Strategies for Ischaemic Syndromes) Registry Investigators. *Lancet* 1998; 352(9127):507-514.

(44) Bleyer AJ, Hylander B, Sudo H, Nomoto Y, de la Torre E, Chen RA et al. An international study of patient compliance with hemodialysis. *JAMA* 1999; 281(13):1211-1213.

FIGURE 1: Summary of the major clinical measures used in SLE trials

SELENA-SLEDAI: Encompasses 24 weighted items scored dichotomously as present or absent in the previous 10 days, thus improvement or worsening of a manifestation is not captured. Overall disease activity is scored over a range of 0 to 105 points. A minimum clinically meaningful score change = a decrease of 6 points (overall improvement) or an increase of 8 points (overall worsening). A designated change in score (≥ 4 points) between baseline and follow up can be used to dichotomise patients into responders or non-responders for overall disease.

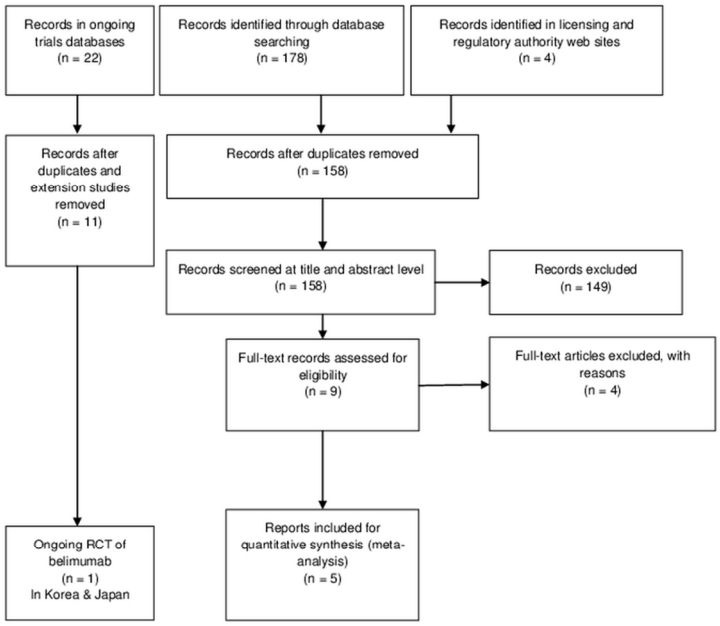
BILAG: Includes 86 items grouped in eight 8 organ systems to assesses organ system involvement over the last four weeks compared to preceding four weeks based on physicians intention to treat using classifications ranging from A to E as follows: A = worsening usually requiring intensification of steroids or immunosuppressant treatments; B = worsening usually requiring antimalarials, nonsteroidal anti-inflammatory drugs (NSAIDs), or low dose steroids; C = stable disease (symptomatic therapy); D = improvement; E = system never involved. Unlike SELENA-SLEDAI it can detect worsening or improvement in individual organ system involvement.

PGA: Is employed to monitor change in patient overall disease activity; typically a visual analogue scale is used ranging between no disease = 0, mild disease = 1, moderate disease = 2, and severe disease = 3.

SRI: A composite instrument (combining elements of SELENA-SLEDAI, BILAG and PGA) developed by belimumab-trialists in conjunction with the US FDA. It allows patients to be dichotomised into responders or non-responders according to predefined assessment criteria in each of the component elements, such as: a SELENA-SLEDAI improvement of ≥ 4 points, plus no worsening in PGA score by > 0.3 points, plus no new BILAG organ system involvement scoring category A in one system or category B in two or more systems. An advantage of SRI, over any one of its components used alone, may be that it can detect SLE improvement in some initial manifestation(s) while guarding against the possibility that worsening in organ systems or overall disease activity might be masked.

90x127mm (300 x 300 DPI)

FIGURE 2: PRISMA 2009 Flow Diagram for Belimumab in SLE RCTs and on-going trials



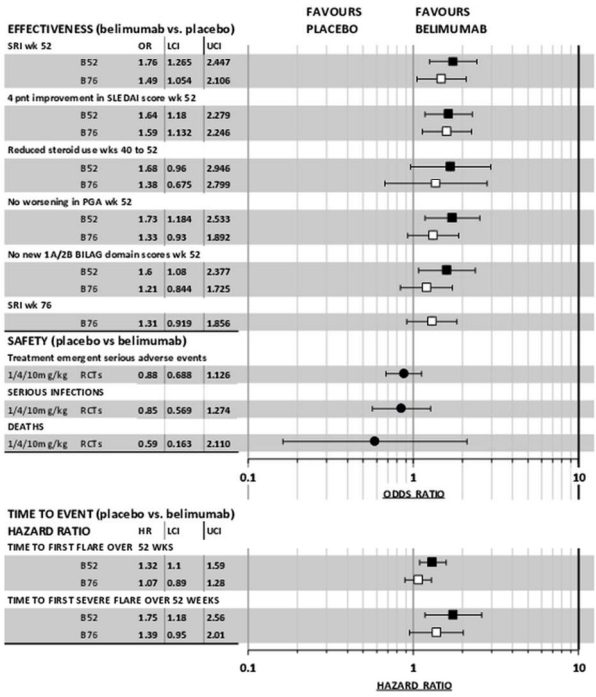
90x127mm (300 x 300 DPI)

FIGURE 3: Differing centre locations in the BLISS 52 and BLISS 76 multicentre trials

123x90mm (300 x 300 DPI)

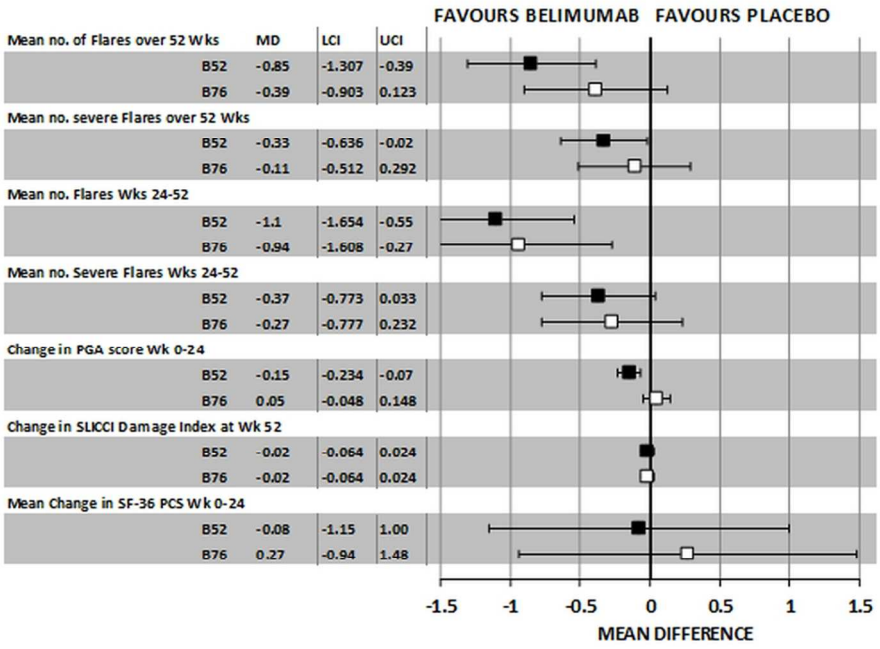
FIGURE 4: Summary of results for major binary and time to event outcomes in belimumab RCTs

Except for safety outcomes the results shown are for the BLISS 52 and BLISS 76 trials. Odds ratios (OR) were calculated from the event rates reported in journal publications; hazard ratios are from data presented in the manufacturer's submission to the FDA. The BLISS trials were well balanced for baseline characteristics (disease, duration, Gender, race, baseline IgG, autoantibody, and complement levels, baseline SLEDAI and PGA scores, BILAG, organ domain involvement, SLICC Damage Index score, and Proteinuria). Safety outcomes are based on data presented in FDA documents.



90x127mm (300 x 300 DPI)

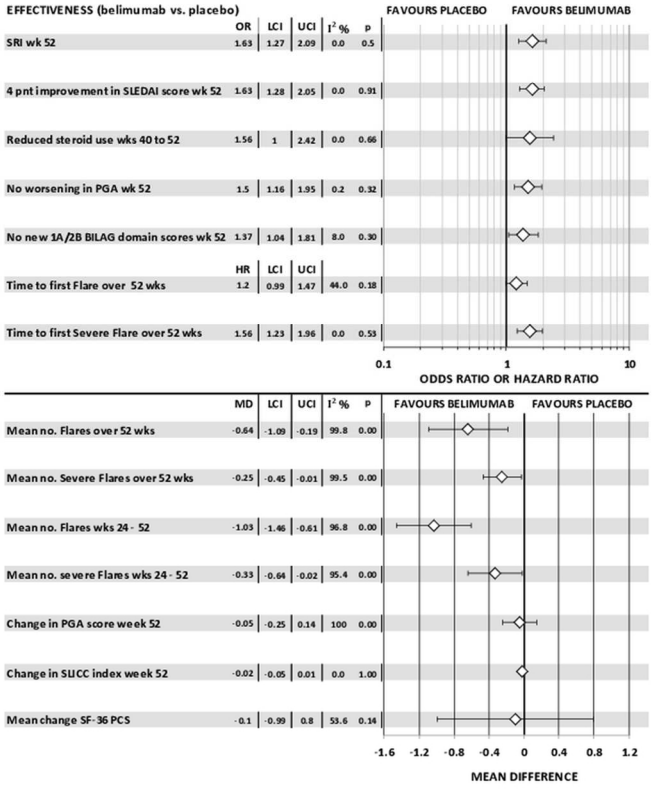
FIGURE 5: Summary of results for major continuous outcomes in BLISS 52 and BLISS 76 trials



107x90mm (300 x 300 DPI)

FIGURE 6: Meta-analysis of major outcomes in the two BLISS trials

Upper panel shows pooled estimates for binary and time to event outcomes (OR = odds ratio; HR = hazard ratio). Lower panel shows pooled estimates for continuous outcomes (MD = mean difference). SLICC = Systemic Lupus International Collaborating Clinics, the SLICC index is a measure of organ damage. Meta-analysis was conducted using random effects method (DerSimonian Laird).



90x127mm (300 x 300 DPI)

Appendix 1

Search Strategies

Cochrane Central Register of Controlled Trials (CENTRAL)

CENTRAL searched via Cochrane Library Interface on 18/05/11

1	MeSH descriptor Lupus Erythematosus, Systemic explode all trees	418
2	(lupus NEAR/3 erythematosus) or (systemic* NEAR/3 lupus) or (SLE)	630
3	(#1 OR #2)	703
4	belimumab OR benlysta	6
5	(#3 AND #4)	4

Medline

Medline searched via Ovid Interface on 19/05/11

1	exp Lupus Erythematosus, Systemic/	42025
2	(lupus adj3 erythematosus).tw.	35497
3	(systemic* adj3 lupus).tw.	31639
4	1 or 2 or 3	50358
5	belimumab.mp.	68
6	benlysta.mp.	3
7	5 or 6	68
8	4 and 7	48
9	randomized controlled trial.pt.	305892
10	controlled clinical trial.pt.	82328
11	randomized.ab.	212836
12	placebo.ab.	124063
13	clinical trials as topic.sh.	153987
14	randomly.ab.	154440
15	trial.ti.	91188
16	9 or 10 or 11 or 12 or 13 or 14 or 15	711420
17	exp animals/ not humans.sh.	3582822
18	16 not 17	656689
19	8 and 18	24

RCT search filter used: Cochrane Highly Sensitive Search Strategy for identifying randomized trials in MEDLINE: sensitivity- and precision-maximizing version (2008 revision); Ovid format. Box 6.4.b in the Cochrane handbook. Reference: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org

Medline In-process

Medline In-Process searched via Ovid Interface on 19/05/11

1	exp Lupus Erythematosus, Systemic/	0
2	(lupus adj3 erythematosus).tw.	1213
3	(systemic* adj3 lupus).tw.	873
4	1 or 2 or 3	1236
5	belimumab.mp.	8
6	benlysta.mp.	4
7	5 or 6	10
8	4 and 7	6

Embase

1	belimumab.mp.orexpbelimumab/	427
2	benlysta.mp.	24
3	1 or 2	428
4	exp systemic lupus erythematosus/	50906
5	(lupus adj3 erythematosus).tw.	40637
6	(systemic: adj3 lupus).tw.	36554
7	4 or 5 or 6	59739
8	3 and 7	302
9	random:.tw.	632763
10	placebo:.mp.	250140
11	double-blind:.tw.	116148
12	9 or 10 or 11	796900
13	8 and 12	144

RCT search filter used: Wong, et al. (2006) Best optimization of sensitivity and specificity. Reference: Wong SS, Wilczynski NL, Haynes RB. Developing optimal search strategies for detecting clinically sound treatment studies in EMBASE. J Med Libr Assoc. 2006 Jan;94(1):41-7. PubMed PMID: 16404468; PubMed Central PMCID: PMC1324770.

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	Page 1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	Page 2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	Page 3-4
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	Page 3-4
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	Exists, available from authors
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	Page 5-6
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	Page 5-6
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Appendix
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	Page 5-6
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	Page 5-6
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	Page 5-6
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	Page 8-10
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	Page 8-12
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	Page 5-6

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	Page 5-6
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	NA
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	Fig 2
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	Page 8-12
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	Page 8-12
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Figures
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	Tables
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	Discussion
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	NA
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	Discussion
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	Discussion
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	Discussion
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	End of paper

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: **www.prisma-statement.org**.