

Belimumab versus anifrolumab in adults with systemic lupus erythematosus: an indirect comparison of clinical response at 52 weeks

Binod Neupane,¹ Pragya Shukla,¹ Mahmoud Slim,^{1,2} Amber Martin,³ Michelle Petri ,⁴ George K Bertsias ,⁵ Alfred H J Kim,⁶ Antonis Fanouriakis ,⁷ Roger A Levy ,⁸ Deven Chauhan,⁹ Nick Ballew ¹⁰

To cite: Neupane B, Shukla P, Slim M, *et al.* Belimumab versus anifrolumab in adults with systemic lupus erythematosus: an indirect comparison of clinical response at 52 weeks. *Lupus Science & Medicine* 2023;**10**:e000907. doi:10.1136/lupus-2023-000907

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/lupus-2023-000907>).

Received 27 January 2023
Accepted 6 April 2023



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Nick Ballew; nick.g.ballew@gsk.com

ABSTRACT

Objective To generate comparative efficacy evidence of belimumab versus anifrolumab in SLE that can inform treatment practices.

Methods The SLE Responder Index (SRI)-4 response at 52 weeks of belimumab versus anifrolumab was evaluated with an indirect treatment comparison. The evidence base consisted of randomised trials that were compiled through a systemic literature review.

A feasibility assessment was performed to comprehensively compare the eligible trials and to determine the most appropriate indirect treatment comparison analysis method. A multilevel network meta-regression (ML-NMR) was implemented that adjusted for differences across trials in four baseline characteristics: SLE Disease Activity Index-2K, anti-double-stranded DNA antibody positive, low complement (C)3 and low C4. Additional analyses were conducted to explore if the results were robust to different sets of baseline characteristics included for adjustment, alternative adjustment methods and changes to the trials included in the evidence base.

Results The ML-NMR included eight trials: five belimumab trials (BLISS-52, BLISS-76, NEA, BLISS-SC, EMBRACE) and three anifrolumab trials (MUSE, TULIP-1, TULIP-2). Belimumab and anifrolumab were comparable in terms of SRI-4 response (OR (95% credible interval), 1.04 (0.74–1.45)), with the direction of the point estimate slightly favouring belimumab. Belimumab had a 0.58 probability of being the more effective treatment. The results were highly consistent across all analysis scenarios.

Conclusions Our results suggest that the SRI-4 response of belimumab and anifrolumab are similar at 52 weeks in the general SLE population, but the level of uncertainty around the point estimate means we cannot rule out the possibility of a clinically meaningful benefit for either treatment. It remains to be seen if specific groups of patients could derive a greater benefit from anifrolumab or from belimumab, and there is certainly an unmet need to identify robust predictors towards more personalised selection of available biological agents in SLE.

WHAT IS ALREADY KNOWN ON THIS TOPIC

- ⇒ Belimumab and anifrolumab are both approved treatments for SLE; the efficacy of each treatment has been demonstrated versus placebo in clinical trials.
- ⇒ The only results to date on the efficacy of belimumab versus anifrolumab are from a single study that indirectly compared the two treatments, and the study had several limitations.

WHAT THIS STUDY ADDS

- ⇒ The clinical response of belimumab and anifrolumab at week 52 is generally comparable, and belimumab has a 0.58 probability of being the more effective treatment.
- ⇒ Our results clearly demonstrate that, despite a recent publication to the contrary, there is no evidence to indicate that patients with SLE would benefit from a change in treatment from belimumab to anifrolumab or vice versa.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

- ⇒ Our results are a valuable reminder for future research that when population-adjusted indirect comparisons are conducted, the patient-level data informing the population adjustment must be large enough and broad enough that the population-adjusted treatment effects can be accurately estimated.

INTRODUCTION

SLE is characterised by chronic inflammation leading to significant morbidity and mortality.¹ Treatment of SLE aims to minimise disease activity, decrease the incidence of disease flares and prevent organ damage.² Conventional treatment options include antimalarials, glucocorticoids and immunosuppressive agents.^{2–3} Although such treatments can be initially successful, patients often require

adjunctive therapies or a switch to different immunosuppressives, including biologic drugs.²

Belimumab, a human immunoglobulin G1 λ (IgG1) monoclonal antibody, inhibits the biologic activity of B-lymphocyte stimulating protein.⁴ Belimumab was first approved in 2011 by the US Food and Drug Administration for patients with active, autoantibody positive SLE receiving standard therapy (ST), and is now approved for the treatment of patients ≥ 5 years of age with SLE in >75 countries.^{4,5} Patients treated with belimumab plus ST have consistently demonstrated a reduction in disease activity, glucocorticoid use and frequency of flares versus placebo plus ST in randomised controlled trials (RCTs).^{6–9}

Anifrolumab, a fully human IgG1K monoclonal antibody that binds to type I interferon (IFN) receptor subunit 1 and inhibits signalling by all type I IFNs, was approved in the USA in 2021 for the treatment of patients with moderate-to-severe SLE receiving ST.^{10,11} Anifrolumab received approval based on evidence across three RCTs, two of which (MUSE¹² and TULIP-2¹¹) showed favourable results versus placebo plus ST, while the primary efficacy endpoint of SLE Responder Index-4 (SRI-4) was not met in the TULIP-1 trial.¹³

In the absence of a head-to-head RCT comparing belimumab and anifrolumab, an indirect treatment comparison (ITC) that incorporates results across the available RCTs can generate robust comparative evidence to inform treatment practices. An ITC across RCTs can produce valid evidence when there are no differences across trials in effect modifiers (EMs), or when differences in EMs are appropriately accounted for.^{14–17} EMs are characteristics that alter the relative effect of a treatment, so that it is more or less effective than an alternative treatment, depending on the level of the EM (further information on ITCs and EMs provided in online supplemental appendix 1). An ITC that adjusts for differences across trials in EMs is referred to as a population-adjusted indirect comparison (PAIC). See online supplemental appendix 2 for more details on PAICs. One PAIC comparing belimumab and anifrolumab has been published;¹⁸ however, the study did not meet the fundamental requirements for a robust population-adjusted analysis.¹⁹ Several studies^{20–22} have demonstrated that PAIC methods can perform poorly and yield inaccurate estimates under scenarios similar to that of the Bruce *et al* study.¹⁸ See Ballew *et al* (and the Discussion section) for further details on the limitations of the Bruce *et al* study.^{18,19}

The primary objective of our study was to generate evidence on the comparative efficacy of the approved doses of belimumab versus anifrolumab at 52 weeks. A secondary objective was to examine the validity of the findings reported in the Bruce *et al* study.¹⁸

METHODS

Compiling and assessing the evidence base

A systematic literature review (SLR) (adhering to the Cochrane Collaboration and Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines) was conducted to identify all trials reporting relevant

outcomes at 52 weeks for adults (≥ 18 years) with SLE receiving belimumab or anifrolumab plus ST, published as of 12 April 2022.^{23,24} A standardised data extraction template was used to capture all relevant information from the included trials (detailed information on the SLR can be found in online supplemental appendix 3).

We identified EMs that would need to be balanced to conduct an unbiased ITC based on clinical knowledge, published literature, an evaluation of reported subgroup results within individual trials, exploratory analyses of individual patient data (IPD) in belimumab trials that GSK had on file and several rounds of discussion with lupus experts. When multiple options were available for how to adjust (eg, adjust for proportion with any glucocorticoid use or proportion with a glucocorticoid dosage threshold), we relied on the exploratory regression analyses to inform our decisions (online supplemental appendix 4).

The evidence base compiled from the SLR was compared in terms of study design, trial circumstances, patient population, treatment implementation and outcome definitions (see online supplemental appendix 5 for details on each trial). Special focus was paid to the comparison of baseline characteristics across trials that were identified as potential EMs.²⁵ Where differences across trials were identified, the expected direction and magnitude of the potential bias was noted, as well as when data limitations precluded a thorough comparison or appropriate adjustment.

Outcomes for analysis

The primary efficacy outcome was proportion of patients achieving SRI-4 response at week 52. In the earlier belimumab trials (BLISS-52, BLISS-76, BLISS-SC and the North East Asia study (NEA)),^{6–9} SRI-4 incorporated Safety of Estrogens in Lupus National Assessment–SLE Disease Activity Index (SELENA-SLEDAI) in the original definition. SRI-4 has since been re-analysed in these trials, incorporating modified scoring for proteinuria adapted from SLEDAI-2K. In EMBRACE, SRI-4 was reported both ways.²⁶ In the anifrolumab trials, SRI-4 incorporated SLEDAI-2K.^{11–13} Thus, in an effort to make as close of a like-for-like comparison as possible, the SRI-4 results that incorporated a modified version of SLEDAI-2K were used from the belimumab trials. However, there were unresolved differences between the measures regarding joint scoring that could not be addressed, as SELENA-SLEDAI requires three joints, but SLEDAI-2K just two joints. Further, there were also differences in SRI-4 in terms of the British Isles Lupus Assessment Group (BILAG) instrument used in the trials (BILAG classic in belimumab trials and BILAG 2004 in anifrolumab trials) that could not be reconciled.

Several other efficacy outcomes were considered for analysis at 52 weeks, including proportion of patients with ≥ 4 -point reduction in SLEDAI, SLEDAI response on specific organ domains, flares, glucocorticoid reduction and anti-double-stranded DNA antibody (anti-dsDNA)

levels. However, robust analyses were not feasible with the evidence base identified (explained in detail in online supplemental appendix 5). Of note, for the proportion of patients achieving ≥ 4 -point reduction in SLEDAI, MUSE was the only anifrolumab trial identified from the SLR that reported this outcome.¹² However, the 4-point reduction in MUSE was based on the clinical components only without consideration of the immunological components. After the SLR was completed, a pooled analysis of TULIP-1 and TULIP-2 ≥ 4 -point reduction in SLEDAI was reported.¹⁸ To ensure the credibility of an ITC, results need to be available for each trial separately (ideally, results are available in terms of the proportion of responders for each arm). Thus, given these data limitations and their potential impact on the credibility of an ITC for ≥ 4 -point reduction in SLEDAI, an ITC of this outcome could not be conducted to provide credible evidence on the comparative efficacy of belimumab versus anifrolumab at 52 weeks. However, exploratory analyses of ≥ 4 -point reduction in SLEDAI were conducted (anifrolumab trial data from the ITC of Bruce *et al*) to aid our understanding of the results of Bruce *et al* for this outcome.¹⁸

Statistical analysis

Analysis scenarios

The feasibility assessment revealed that there were differences in EMs for SRI-4 between the belimumab and anifrolumab studies, meaning PAIC methods would be required to conduct an unbiased ITC of SRI-4. For the primary outcome of SRI-4, a fixed effects (FE) multilevel network meta-regression (ML-NMR) model that adjusted for all possible 'imbalanced EMs', but no prognostic variables, was selected as the base-case. The rationale for this decision was that the model was capable of adjusting for any meaningful bias introduced by EMs, without making any sacrifices in terms of simplifying the network structure. Adjusting for all possible EMs (specifically Black African ancestry) would have required pooling some belimumab trials together and treating them as a single trial. Four separate sensitivity analyses were conducted for SRI-4 to assess the robustness of our results to alternative sets of variables for adjustment and alternative PAIC methods (simulated treatment comparison (STC) and matching adjusted indirect comparison (MAIC)) methods. Of note, ML-NMR and STC are similar in that, in addition to including EMs in the model (to account for bias), prognostic variables can also be included (no interaction with treatment) to obtain more precise estimates. This is in contrast to MAIC, where only EMs should be included.

Additional sensitivity analyses were also conducted to understand the impact of using the modified SRI-4 definition for the belimumab trials (sensitivity analysis used the SRI-4 definition for the belimumab trials based on SELENA-SLEDAI) and to understand the impact of treating belimumab intravenous and subcutaneous formulations as equivalent treatments (sensitivity with

altered network structure, so belimumab intravenous and subcutaneous were individual treatment nodes and compared with anifrolumab separately). Additional details on the analyses can be found in online supplemental appendix 6. Of note, Sensitivity 3 (described in online supplemental appendix 6) was the preplanned base-case analysis but was moved to a sensitivity because it required treating the five available belimumab trials as three trials (BLISS-52 and BLISS-76 were pooled and the NEA study and EMBRACE were pooled). Relatedly, STC and MAIC methods suffer from a similar limitation in that they can only be applied to simple networks of evidence, and as a result, we had to pool all belimumab trials together and all anifrolumab trials together in STC and MAIC analyses.

Exploratory analyses were conducted to emulate the approach implemented in Bruce *et al* for the clinical response outcomes SRI-4 and ≥ 4 -point reduction in SLEDAI.¹⁸ Specifically, this meant conducting the analyses with the same evidence base and with the same results for the anifrolumab trials as reported in Bruce *et al*,¹⁸ which in some cases differed from the results previously published for the trials. We also used the same methods (STC and MAIC), network structure and set of EMs as in Bruce *et al*.¹⁸ All exploratory analyses were undertaken as in Bruce *et al*,¹⁸ with the original SRI-4 and ≥ 4 -point reduction in SLEDAI definitions for the belimumab trials based on SELENA-SLEDAI and the SRI-4 definition for the anifrolumab trials incorporating SLEDAI-2K. However, IPD from the belimumab trials were used to inform the population adjustments, instead of IPD from the anifrolumab trials as in Bruce *et al*.¹⁸ Importantly, the IPD from the belimumab trials is a larger sample than that from the anifrolumab trials (1125 vs 710) and is representative of a broader SLE population (includes patients with and without BILAG ≥ 1 A or ≥ 2 B at baseline).

Model implementation

The steps in ML-NMR include deriving the aggregate level likelihood and then deriving the integral in the aggregate model. Deriving the aggregate-level model in ML-NMR requires using IPD from the trials to inform the covariate distributions and correlation structure of variables from the studies. While IPD was available (and used) for the belimumab trials, the IPD for the anifrolumab trials was not. Thus, the observed distributions and correlations from the belimumab trials were used to inform the distributions and correlations in the anifrolumab trials. The FE model used a non-informative normal prior distribution (location=0, scale=100) on each parameter of interest. Three chains (7000 iterations, out of which the first 4000 were the burn-in iterations) were run on each ML-NMR. A random effects (RE) ML-NMR (half-normal (location=0, scale=0.5) prior distribution for the between-study SD) was also conducted for each FE ML-NMR as a check for residual heterogeneity remaining after adjusting for the selected EMs. ML-NMR was implemented in a Bayesian framework by using Markov chain

Monte Carlo sampling and with the ‘multinma’ package in R.²⁷ Median ORs and 95% credible intervals (CrI) were reported. Treatment-rank probabilities were produced, as well as surface under the cumulative ranking curve (SUCRA) values. The relative effects, ranking probabilities and SUCRA values were estimated for each study population of interest (each individual trial population included in the network, as well as the combined anifrolumab and belimumab populations).

As noted above, for our MAIC and STC analyses, we had to pool all belimumab trials together and all anifrolumab trials together and treat them as two large pseudo-trials. The MAIC and STC analyses were then conducted following the methods described by Signorovitch *et al* and National Institute for Health and Care Excellence guidelines.^{28 29} See online supplemental appendix 2 for full model implementation details.

Patient and public involvement

Patients or the public were not involved in the design, or conduct, or reporting or dissemination plans of our research.

RESULTS

Clinical response of belimumab versus anifrolumab at 52 weeks

Evidence base

Nineteen unique trials were identified by the SLR. The detailed findings of the SLR and feasibility assessment are included in online supplemental appendix 5. Eight of the 19 trials that were identified were ultimately eligible for the SRI-4 analysis at 52 weeks comparing the approved doses of belimumab (10 mg/kg intravenous and 200 mg subcutaneous) and anifrolumab (300 mg intravenous): BLISS-52 (NCT00424476); BLISS-76 (NCT00410384); BLISS-SC (NCT01484496); NEA study

(NCT01345253); EMBRACE (NCT01632241); TULIP-1 (NCT02446912); TULIP-2 (NCT02446899); MUSE (NCT01438489).^{6-9 11-13 26} The trial-level SRI-4 results are presented in figure 1. More detailed information on the inclusion criteria, intervention, baseline characteristics and outcome definitions for these trials is included in online supplemental appendix 5.

ITC SRI-4 at 52 weeks

Eight characteristics were identified as likely EMs for SRI-4 (table 1). Accordingly, the trials would need to be balanced in terms of these characteristics to conduct an unbiased ITC. However, data limitations precluded the possibility of evaluating (and potentially adjusting) the level of balance for two of the variables; body mass index (BMI) was not available in MUSE and none of the trials reported smoking status. Thus, it was possible to adjust for six (SLEDAI-2K, Black African ancestry, low C3, low C4, anti-dsDNA and any glucocorticoid use) of the potential eight EMs. For two (Black African ancestry and any glucocorticoid use) of these six EMs, the level of imbalance was negligible (table 1). Of the remaining four, if no population adjustment was made, one of the variables would be expected to introduce bias in favour of anifrolumab (SLEDAI-2K) and three would be expected to introduce bias in favour of belimumab (low C3, low C4, anti-dsDNA).

In the base-case ML-NMR analysis of the SRI-4 outcome that adjusted for the four imbalanced EMs, belimumab and anifrolumab were generally comparable, with the direction of the point estimate slightly favouring belimumab (OR (95% CrI) 1.04 (0.74–1.45)). There was a 0.58 probability that belimumab was the more effective treatment and a 0.42 respective probability for anifrolumab. Of note, while the model predictions were in line with the observed SRI-4 results in the belimumab trials, the

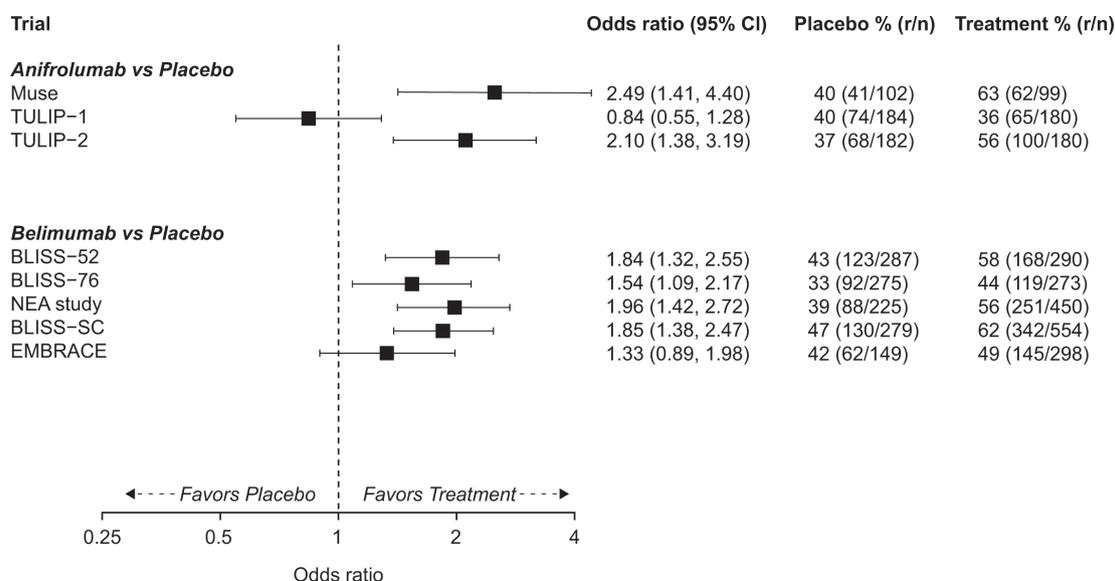


Figure 1 Trial level results that contributed to the ITC for SRI-4 at 52 weeks. ITC, indirect treatment comparison; n, sample size; r, number of responders; SRI-4, SLE Responder Index-4.

Table 1 Potential treatment EMs for SRI-4: characteristics that need to be balanced across trials

Baseline characteristic	Expected effect-modifying relationship (relative to placebo) with outcome	Anticipated bias (direction and magnitude) for indirect comparison*	Rationale for anticipated bias (values are belimumab trials vs anifrolumab trials)	Adjustment in model	
SLEDAI-2K	Treatment effect is larger in populations with higher SLEDAI-2K values		Moderate/large bias in favour of anifrolumab	Strong treatment EM with moderate difference (mean of 10.4 vs 11.3)	Balance mean SLEDAI-2K across trials
Race	Treatment effect is smaller in Black African ancestry race than other races		Negligible bias in favour of anifrolumab	Moderate treatment EM with small difference (20% vs 14% Black African Ancestry)	Not adjusted in the base-case analysis†
C3	Treatment effect is larger among patients with low C3 concentration		Moderate bias in favour of belimumab	Moderate treatment EM with moderate difference (49% vs 36% low C3)	Balance proportion with low C3 across trials
C4	Treatment effect is larger among patients with low C4 concentration		Moderate bias in favour of belimumab	Moderate treatment EM with moderate difference (37% vs 23% low C4)	Balance proportion with low C4 across trials
Anti-dsDNA	Treatment effect is larger among anti-dsDNA positive patients		Small bias in favour of belimumab	Small treatment EM with large difference (71% vs 40% positive)	Balance proportion anti-dsDNA positive‡
Glucocorticoid use	Treatment effect is larger among patients with any glucocorticoid use		Negligible bias in favour of belimumab	Moderate treatment EM with small difference (88% vs 82% with any use)	Not adjusted in the base-case analysis†
Smoking status	Treatment effect is smaller in smokers	?	Bias for characteristic is unknown	Smoking status is not reported for any trial in the evidence base	None
BMI	Treatment effect is smaller in patients with high BMI	?	Bias for characteristic is unknown	BMI is only available for some trials (but appears similar across trials that report it)§	None

*Arrow thickness indicates the strength of the effect.
†Characteristic was adjusted for in sensitivity analyses.
‡Positivity in belimumab trials defined based on 30 IU/mL threshold, while positivity in anifrolumab trials was defined based on 15 IU/mL threshold. While it was possible to alter the definition in the belimumab trials to match the definition in the anifrolumab trials, a 15 IU/mL threshold was not clinically meaningful for the belimumab trials.
§BMI of 27.6 kg/m² for pooled TULIP trials (reported in Bruce *et al*¹⁶) and BMI of 25.4 kg/m² for pooled belimumab trials.
Anti-dsDNA, anti-double-stranded DNA antibody; BMI, body mass index; C3/C4, complement component 3/4; EM, effect modifier; SLEDAI-2K, SLE Disease Activity Index 2000; SRI-4, SLE Responder Index-4.

predictions for the anifrolumab trials did not follow the observed study-level SRI-4 results for the three anifrolumab trials (based on visual comparison of observed trial-level results in [figure 1](#) and model predictions in [figure 2](#)). To this point, the deviance information criterion from the RE model was only marginally lower than the base-case FE model (4076 vs 4078), indicating similar model fit. However, the estimate for the heterogeneity parameter was relatively large ($\tau=0.26$) and was accompanied by a relatively large amount of uncertainty (SD of $\tau=0.15$).

The ORs of belimumab versus anifrolumab were highly consistent between the base-case and all sensitivity analyses (sensitivity analyses that employed alternative sets of variables for adjustment and alternative PAIC methods in [figure 3](#); additional analysis results can be found in online supplemental appendix 6). The base-case and sensitivity

analysis results were also in line with the results of the standard FE Bayesian network meta-analysis (NMA; OR (95% CrI) 1.13 (0.83–1.53)). Convergence to the posterior distribution was achieved in all Bayesian (NMA and ML-NMR) analyses.

Emulating the approach of Bruce *et al*

The results obtained for SRI-4 when emulating the Bruce *et al*¹⁸ approach suggested that belimumab and anifrolumab were generally comparable, with the direction of the point estimate slightly favouring belimumab (STC OR (95% CI) 1.06 (0.65 to 1.72); MAIC OR (95% CI) 1.11 (0.66 to 1.86)).

The results from the two exploratory analyses with ≥ 4 -point reduction in SLEDAI also suggested that belimumab and anifrolumab were generally comparable, with

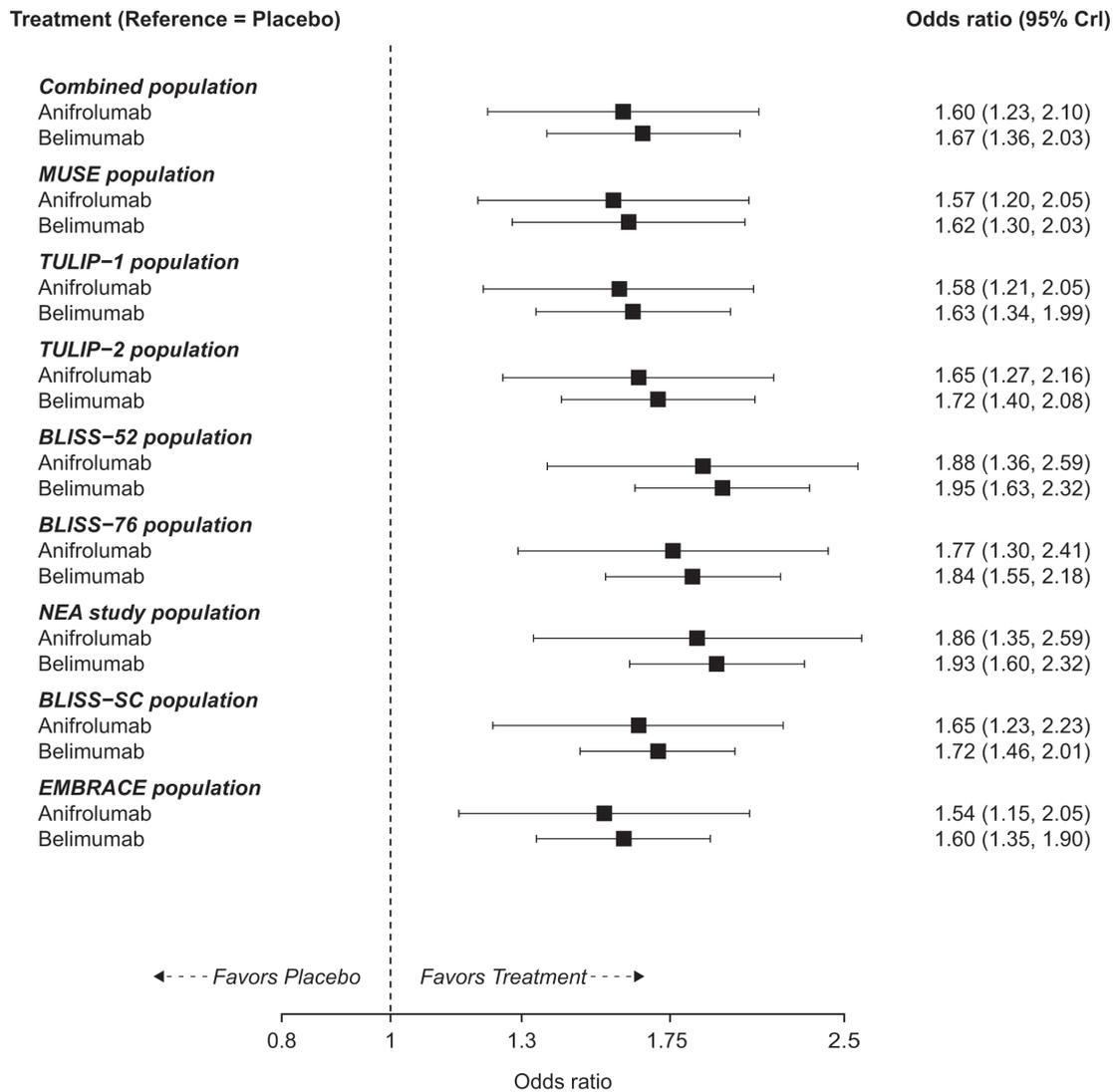


Figure 2 Predicted ORs for belimumab plus standard therapy and anifrolumab plus standard therapy versus placebo plus standard therapy for the base-case ML-NMR analysis of SRI-4 at 52 weeks in each population. Combined population is the pooled population across the three anifrolumab trials. CrI, credible interval; ML-NMR, multilevel network meta-regression; SRI-4, SLE Responder Index-4.

the direction of the point estimate slightly favouring belimumab (STC OR (95% CI) 1.15 (0.71 to 1.86); MAIC OR (95% CI) 1.14 (0.68 to 1.92)).

DISCUSSION

This study implemented a PAIC of RCT data to evaluate the efficacy of belimumab versus anifrolumab at 52 weeks in adults with SLE. The results of our analysis suggest that belimumab and anifrolumab are generally comparable in terms of SRI-4 at 52 weeks, but we cannot rule out the possibility of a clinically meaningful benefit for either treatment. Our results were consistent across the host of sensitivity analyses conducted. Given the differences identified in potential EMs, the ML-NMR results (and results from our other population adjustment models) are assumed to be less biased than the results using standard Bayesian NMA. Nonetheless, much of the bias in a standard Bayesian NMA appears to cancel

out (some in favour of anifrolumab and some in favour of belimumab), so the results of the Bayesian NMA and PAIC analyses are largely consistent.

A key requirement of ITCs is that either the populations are inherently similar in terms of EMs (in the case of a standard ITC), or in the case of a PAIC, that they are appropriately adjusted to remove any inherent differences so that unbiased estimates can be obtained. When population adjustments are necessary, the population sample contributing the IPD must be large enough and broad enough to accurately estimate the treatment effects in the comparator population.^{16 20} Our primary analyses with SRI-4 clearly met this requirement, with the IPD population sample (the five belimumab trials) consisting of >3000 patients, which was broad enough to accurately estimate the treatment effects in the anifrolumab population. The total sample size of our IPD in the MAIC of SRI-4 was 3080, with an effective sample size (ESS)

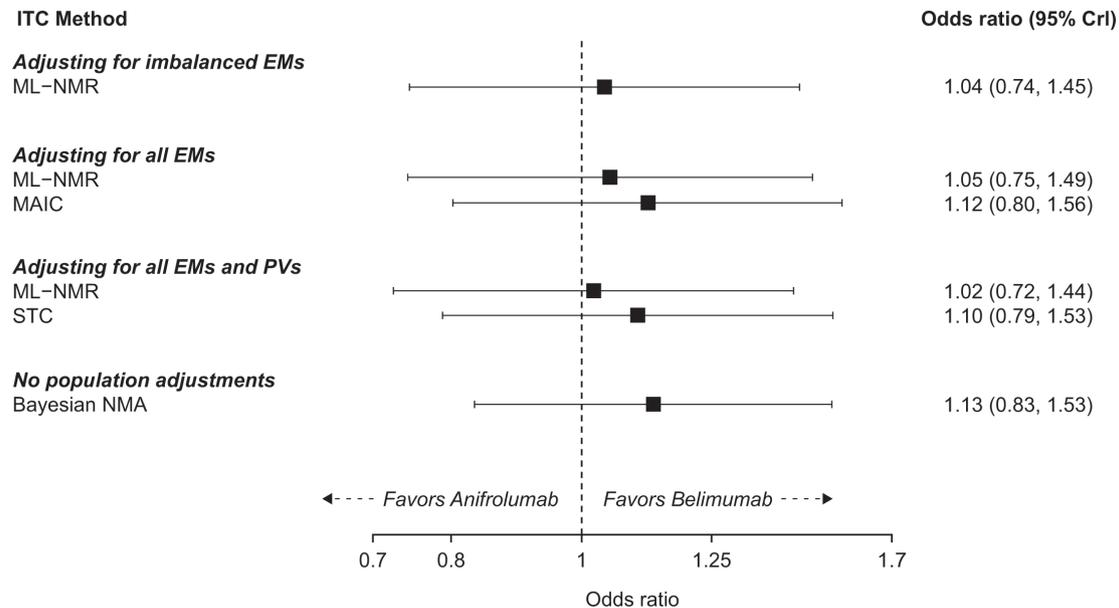


Figure 3 SRI-4 results at 52 weeks of belimumab plus standard therapy versus anifrolumab plus standard therapy for the base-case and sensitivity analyses. CrI, credible interval; EM, effect modifier; ITC, indirect treatment comparison; MAIC, matching-adjusted indirect comparison; ML-NMR, multilevel network meta-regression; NMA, network meta-analysis; PV, prognostic variable; SRI-4, SLE Responder Index-4; STC, simulated treatment comparison.

post-adjustment of 1531. While this represents a sizeable reduction from the full sample size used to estimate the OR of belimumab versus placebo across the five belimumab trials, it is still a robust sample size to use for a PAIC.

An unexpected finding from our extensive set of PAIC analyses was that the high level of heterogeneity in SRI-4 results at 52 weeks across the three anifrolumab trials appears largely unrelated to any population differences across these three trials in EMs. This is true not only for our set of EMs but also appears to be true for the more extensive set that was identified in Bruce *et al.*¹⁸ Consequently, our population-adjusted analyses can successfully explain the observed variation in SRI-4 for the belimumab trials, but not the differences in the trial-level SRI-4 results for the three anifrolumab trials. This finding highlights the fact that more research on anifrolumab is needed (which is beyond the scope of this study) to fully understand the effect of anifrolumab on SRI-4 at 52 weeks in the general SLE population. In the context of the current study, this finding means that the level of uncertainty around the placebo versus anifrolumab comparison and around the belimumab versus anifrolumab comparison may be even larger than what is estimated in our population-adjusted models.

When emulating the Bruce *et al.*¹⁸ approach, we obtained estimates in line with our primary analyses. These results are significantly different from those reported in Bruce *et al.*¹⁸ For example, whereas we obtained an SRI-4 OR of 1.11 with the MAIC, Bruce *et al.*¹⁸ obtained an OR of 0.34 (reported as 2.91 in their publication as belimumab was used as the reference treatment). The key difference between our emulation of the Bruce *et al.* approach and the actual approach in Bruce *et al.* is that we had access to different IPD (we used IPD from the belimumab trials

and Bruce *et al.* used IPD from anifrolumab trials).¹⁸ Consequently, our comparison was made in the combined anifrolumab trial population and the Bruce *et al.*¹⁸ comparison was made in the combined belimumab trial population (MAICs and STCs estimates can only be produced within the population that does not have IPD). If belimumab and anifrolumab treatment effects were modified in entirely different ways by the EMs, then it would be theoretically possible for both results to be correct. However, this is not considered clinically plausible, and therefore, other explanations are more likely.

It is likely that most or all of the differences between our results when emulating the Bruce *et al.* approach and results in Bruce *et al.* can be explained by the fact that Bruce *et al.* did not have sufficient IPD available to undertake their approach.¹⁸ As reported in Bruce *et al.*¹⁸ the total sample size from the two anifrolumab trials (TULIP-1 and TULIP-2) in the MAIC of SRI-4 was 710 and the ESS post-adjustment of these two trials was only 71 patients (a 90% reduction). This can be loosely interpreted to mean that only 71 patients were used to inform the anifrolumab versus placebo comparison that was indirectly compared with belimumab. In contrast, the total sample size of our IPD in the MAIC of SRI-4 when emulating the Bruce *et al.*¹⁸ approach was 1125 (BLISS-52 and BLISS-76), with an ESS post-adjustment of 351 (approximately a 69% reduction). Thus, when emulating the Bruce *et al.* approach,¹⁸ we had an ESS approximately five times the size of what was available to inform the population adjustment produced in Bruce *et al.* When comparing the ESS from our primary analysis (n=1531) to that of Bruce *et al.* (n=71),¹⁸ our ESS is over 20 times larger.

It is also important to note that, beyond having limited IPD, there are further limitations to the Bruce *et al.*

approach.^{18 30} First, not all eligible trials in the evidence base were included in the analysis. Bruce *et al* contend that this was a necessary limitation due to issues with how STC and MAIC methods must be implemented.³¹ However, the ML-NMR method we used does not suffer from the issues they allude to.³¹ ML-NMR can be incorporated for any connected network of evidence and also provides a way to check assumptions (via a RE model) and evaluate model performance.^{17 32} Thus, as we have demonstrated here, there is no need to remove eligible trials from the evidence base. Second, Bruce *et al*¹⁸ employed SRI-4 values for the TULIP trials (≥ 4 -point reduction in SLEDAI has not been reported elsewhere so could not be verified) that were higher than previously reported in the primary TULIP publications: OR of 1.63 reported in Bruce *et al*, while an OR of 1.33 would be expected based on a pooling of the prespecified results in the primary publications.^{11 13} One possible explanation for the discrepancy could be that Bruce *et al* employed results from a post hoc analysis of the TULIP-1 SRI-4 results.¹⁸ However, even if the revised post hoc definition for TULIP-1 was used when pooling the trials, the OR would be 1.56. Third, Bruce *et al*¹⁸ adjusted for the proportion of patients with BILAG ≥ 1 A or ≥ 2 B at baseline in the trials, despite the belimumab and anifrolumab trials using different versions of the BILAG (belimumab trials used the BILAG Classic; the anifrolumab trials used the BILAG 2004). In particular, the BILAG 2004 added two new organs/systems, removed the vasculitis section and rearranged other organ systems.³³ Thus, the apparent differences in BILAG across the trial populations may just be an artefact of the different instruments. This issue is further compounded because the apparent difference in proportion of patients with BILAG ≥ 1 A or ≥ 2 B at baseline in the belimumab and anifrolumab trials appears to be the primary driver of why the IPD sample of Bruce *et al*¹⁸ had poor overlap with the belimumab trial population. There were only approximately 40 patients (5.6% of the sample) in the anifrolumab trials that had no BILAG ≥ 1 A or ≥ 2 B, and yet these 40 patients would have needed to account for 39% of the sample in order to align with the belimumab trial population. With such a small group of patients, even altering the results of just two or three patients (eg, observing 4 of 20 responses vs 7 out of 20 responses in the placebo arm) could have a dramatic impact on the overall results.

Our study also had limitations, mainly that our efficacy analyses were limited to a single outcome (SRI-4) and could only be conducted at 52 weeks. While SRI-4 has been associated with improvements in clinical, laboratory and patient-reported outcome measures,^{34 35} no single outcome provides a comprehensive view of efficacy. With SRI-4, SLEDAI is used to assess improvement, while BILAG and Physician Global Assessment are incorporated to capture worsening. Thus, analyses of SRI-4 alongside outcomes that assess improvement in terms of BILAG (such as BILAG-based Composite Lupus Assessment) would provide a more nuanced picture of

belimumab's potential to improve disease activity relative to anifrolumab. Similarly, Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index, which is a key measure for disease modification in SLE, represents another important dimension of efficacy not considered in this study.³⁶ Further analyses at other timepoints would be valuable to better understand how quickly both treatments become effective and how well efficacy is maintained. Although we assessed the feasibility of numerous other efficacy analyses, it was not possible to undertake analyses with any other endpoints. Even within our analysis of SRI-4, it must be noted that there were differences across trials in the precise definition of SRI-4 that was employed. Specifically, there were potential differences in joint scoring for the SLEDAI component of SRI-4 and there were also differences in terms of the BILAG instrument incorporated in the trials.

Our SRI-4 analyses were also unable to adjust for all eight of the EMs identified (adjusted for four in the base-case and six in the sensitivity). Specifically, none of our analyses adjusted for BMI or smoking status. Thus, it is possible that there was residual confounding in our analysis due to differences in BMI and smoking status across the belimumab and anifrolumab trials. We believe this is unlikely for BMI based on the limited BMI information that is available. However, the magnitude of the difference in the proportion of smokers is unknown. Beyond the EMs that could not be accounted for, there are also differences in time periods that the trials span (the anifrolumab trials were conducted in a post-belimumab world), which may translate into important differences in ST and prior therapies received at baseline. While the methodology we have used (only comparing the ORs across trials as opposed to the absolute proportion of responders) should mostly protect our results from being affected by this issue, we acknowledge the potential that some differences could still modify the treatment effects. A further limitation is that we only had access to the belimumab trial IPD and consequently had to make the 'shared EMs' assumption (that anifrolumab vs placebo relative effects are modified in the same way as belimumab vs placebo) to conduct the ML-NMR. If this assumption is violated, the results of the ML-NMR may be called into question.³² However, the results of the ML-NMR, STC and MAIC are all very consistent and the latter two methods do not explicitly require the shared EM assumption (even when the shared EM assumption is violated, STC and MAIC are still unbiased in the specific population in which the analysis was undertaken). Thus, at worst case, the results may not be generalisable to other SLE populations. The fact that we did not have access to the anifrolumab IPD also meant we had to assume the type of marginal distribution of covariates and the correlation structure for the anifrolumab trials (not reported in the anifrolumab trials) based on what was observed in the belimumab trials.

In conclusion, we performed a robust PAIC analysis that suggests belimumab and anifrolumab are generally

comparable in terms of SRI-4 response at 52 weeks. Future comparisons of belimumab versus anifrolumab may be valuable as more data for anifrolumab become available. It remains to be seen if specific groups of patients could derive a greater benefit from anifrolumab or from belimumab, and there is certainly an unmet need to identify robust predictors towards more personalised selection of available biological agents in SLE. However, our study did not find evidence to support that patients with SLE as a group would benefit from a change in treatment practices from belimumab to anifrolumab or vice versa.

Author affiliations

- ¹Evidence Synthesis, Modeling & Simulation, Evidera, St-Laurent, Quebec, Canada
²Institute of Neurosciences “Federico Olóriz”, University of Granada, Granada, Spain
³Evidence Synthesis, Modeling & Communication, Evidera, Waltham, Massachusetts, USA
⁴Division of Rheumatology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
⁵Department of Rheumatology, Clinical Immunology and Allergy, University of Crete School of Medicine, Crete, Greece
⁶Division of Rheumatology, Department of Medicine, Washington University School of Medicine, Saint Louis, Missouri, USA
⁷First Department of Propaedeutic Internal Medicine, “Laikon” General Hospital, National Kapodistrian University of Athens Medical School, Athens, Greece
⁸Specialty Care, Global Medical Affairs, GSK, Collegeville, Pennsylvania, USA
⁹Value Evidence and Outcomes, GSK, Brentford, UK
¹⁰Value Evidence and Outcomes, GSK, Collegeville, Pennsylvania, USA

Acknowledgements The authors would like to thank Kerry Gairy, previously an employee of GSK, and Kyle Fahrback, Evidera, for their contributions to the development of the study’s protocol and data interpretation. Medical writing support was provided by Helen Taylor, PhD, Fishawack Indicia Ltd, UK, part of Fishawack Health, and was funded by GSK.

Contributors BN, PS, MS and AM contributed to the acquisition of the data. BN, PS, MS, AM, RAL, DC and NB contributed to the conception or design of the study. All authors contributed to the analysis or interpretation of the data. NB is the guarantor of the manuscript.

Funding This study was funded by GSK (GSK Study 214174).

Competing interests BN, PS, MS and AM are employees of Evidera, a part of Thermo Fisher Scientific. MP has received grant/research support from AstraZeneca, Aurinia, Eli Lilly, Exagen, GSK, Janssen and Thermo Fisher. MP has also received consulting fees from the BPR Scientific Advisory Committee, Alexion, Amgen, AnaptysBio, Argenx, AstraZeneca, Aurinia, AxDev, Biogen, Boston Pharmaceuticals, Caribou Biosciences, CVS Health, Eli Lilly, Gilead Biosciences, GSK, Idorsia Pharmaceuticals, Janssen, Kezar Life Sciences, Kira Pharmaceuticals, Momenta Pharmaceuticals, Nimbus Lakshmi, Proviant, Sanofi, SinoMab and UCB. MP has received speakers’ fees from Aurinia, MedShr and Arthros-FocusMedEd, and has received consulting fees for participation in a data safety monitoring board or advisory board for EMD Serono, Emergent Biosolutions, IQVIA and PPD Development. GKB has received consulting fees from Pfizer, Lilly and Novartis, and has received honorary fees from GSK, AstraZeneca, Pfizer, Novartis, Aenorasis, AbbVie and Lilly. GKB has also received a research grant from Pfizer. AHJK has received research support to Washington University from the National Institute of Arthritis and Musculoskeletal and Skin Diseases (grant number P30 AR073752), National Center for Advancing Translational Sciences (grant number UL1 TR002345), Leona M. and Harry B. Helmsley Charitable Trust, Rheumatology Research Foundation, and National Multiple Sclerosis Society, GSK, and Foghorn Therapeutics. AHJK has performed consultancy for Alexion Pharmaceuticals, ANI Pharmaceuticals, AstraZeneca, Aurinia Pharmaceuticals, Exagen Diagnostics, GSK, Kypha and Pfizer unrelated to this work. AHJK has received payment or honoraria (for lectures, presentations, speakers bureaus, manuscript writing or educational events) from AstraZeneca, Aurinia Pharmaceuticals, Exagen Diagnostics and GSK. AHJK has participated on a data safety monitoring board or advisory board for National Institutes of Health/National Institute of Arthritis and Musculoskeletal and Skin Diseases. AHJK has been a board member for the Rheumatology Research Foundation Scientific Advisory Board and the Lupus Foundation of America-

Heartland Chapter, and president of the St Louis Rheumatology Association. AHJK is the inventor of patent number 11029318 with Kypha unrelated to this work. The funders had no role in the decision to publish or preparation of this manuscript. The content is solely the responsibility of the authors and does not necessarily represent the official views of Washington University, its affiliated academic health care centers, or the National Institutes of Health. AF has received honoraria and consulting fees from GSK, Aenorasis and AstraZeneca. AF has been a paid speaker for AbbVie, Amgen, Pfizer, Lilly, Genesis-Pharma, Novartis, UCB and Boehringer-Ingelheim. RAL, DC and NB are employees of GSK and hold stocks and shares in GSK.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request.

Anonymised individual participant data for belimumab trials and the study documents for this analysis can be requested for further research from www.clinicalstudydatarequest.com.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Michelle Petri <http://orcid.org/0000-0003-1441-5373>
 George K Bertias <http://orcid.org/0000-0001-5299-1406>
 Antonis Fanouriakis <http://orcid.org/0000-0003-2696-031X>
 Roger A Levy <http://orcid.org/0000-0001-6393-6031>
 Nick Ballew <http://orcid.org/0000-0003-3507-0028>

REFERENCES

- Kaul A, Gordon C, Crow MK, *et al*. Systemic lupus erythematosus. *Nat Rev Dis Primers* 2016;2:16039.
- Fanouriakis A, Kostopoulou M, Alunno A, *et al*. 2019 update of the EULAR recommendations for the management of systemic lupus erythematosus. *Ann Rheum Dis* 2019;78:736–45.
- Hill DD, Eudy AM, Egger PJ, *et al*. Impact of systemic lupus erythematosus disease activity, hydroxychloroquine and NSAID on the risk of subsequent organ system damage and death: analysis in a single US medical centre. *Lupus Sci Med* 2021;8:e000446.
- GSK. Benlysta us prescribing information. Available: https://gskpro.com/content/dam/global/hcpportal/en_US/Prescribing_Information/Benlysta/pdf/BENLYSTA-PI-MG-IFU.PDF [Accessed Jul 2022].
- Levy RA, Gonzalez-Rivera T, Khamashta M, *et al*. 10 years of belimumab experience: what have we learnt? *Lupus* 2021;30:1705–21.
- Navarra SV, Guzmán 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;377:721–31.
- 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;63:3918–30.
- Stohl W, Schwarting A, Okada M, *et al*. Efficacy and safety of subcutaneous belimumab in systemic lupus erythematosus: a fifty-two-week randomized, double-blind, placebo-controlled study. *Arthritis Rheumatol* 2017;69:1016–27.
- Zhang F, Bae S-C, Bass D, *et al*. A pivotal phase III, randomised, placebo-controlled study of belimumab in patients with systemic

- lupus erythematosus located in China, Japan and South Korea. *Ann Rheum Dis* 2018;77:355–63.
- 10 Riggs JM, Hanna RN, Rajan B, *et al*. Characterisation of anifrolumab, a fully human anti-interferon receptor antagonist antibody for the treatment of systemic lupus erythematosus. *Lupus Sci Med* 2018;5:e000261.
 - 11 Morand EF, Furie R, Tanaka Y, *et al*. Trial of anifrolumab in active systemic lupus erythematosus. *N Engl J Med* 2020;382:211–21.
 - 12 Furie R, Khamashta M, Merrill JT, *et al*. Anifrolumab, an anti-interferon- α receptor monoclonal antibody, in moderate-to-severe systemic lupus erythematosus. *Arthritis & Rheumatology* 2017;69:376–86.
 - 13 Furie RA, Morand EF, Bruce IN, *et al*. Type I interferon inhibitor anifrolumab in active systemic lupus erythematosus (TULIP-1): a randomised, controlled, phase 3 trial. *The Lancet Rheumatology* 2019;1:e208–19.
 - 14 Dias S, Sutton AJ, Ades AE, *et al*. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Med Decis Making* 2013;33:607–17.
 - 15 Salanti G, Higgins JPT, Ades AE, *et al*. Evaluation of networks of randomized trials. *Stat Methods Med Res* 2008;17:279–301.
 - 16 Phillippo DM, Ades AE, Dias S, *et al*. Methods for population-adjusted indirect comparisons in health technology appraisal. *Med Decis Making* 2018;38:200–11.
 - 17 Phillippo DM, Dias S, Ades AE, *et al*. Multilevel network meta-regression for population-adjusted treatment comparisons. *J R Stat Soc Ser A Stat Soc* 2020;183:1189–210.
 - 18 Bruce IN, Golam S, Steenkamp J, *et al*. Indirect treatment comparison of anifrolumab efficacy versus belimumab in adults with systemic lupus erythematosus. *J Comp Eff Res* 2022;11:765–77.
 - 19 Ballew N, Mian A, Levy RA, *et al*. Letter to the editor: indirect treatment comparison of anifrolumab efficacy versus belimumab in adults with systemic lupus erythematosus. *J Comp Eff Res* 2023;12:e220106.
 - 20 Phillippo DM, Dias S, Ades AE, *et al*. Assessing the performance of population adjustment methods for anchored indirect comparisons: a simulation study. *Stat Med* 2020;39:4885–911.
 - 21 Remiro-Azócar A, Heath A, Baio G. Methods for population adjustment with limited access to individual patient data: a review and simulation study. *Res Synth Methods* 2021;12:750–75.
 - 22 Remiro-Azócar A, Heath A, Baio G. Parametric G-computation for compatible indirect treatment comparisons with limited individual patient data. *Res Synth Methods* 2022;13:716–44.
 - 23 Cochrane. Cochrane Handbook for systematic reviews of interventions. Available: <https://training.cochrane.org/handbook/current> [Accessed Oct 2022].
 - 24 Page MJ, McKenzie JE, Bossuyt PM, *et al*. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71.
 - 25 Cope S, Zhang J, Saletan S, *et al*. A process for assessing the feasibility of a network meta-analysis: a case study of everolimus in combination with hormonal therapy versus chemotherapy for advanced breast cancer. *BMC Med* 2014;12:93.
 - 26 Ginzler E, Guedes Barbosa LS, D’Cruz D, *et al*. Phase III/IV, randomized, fifty-two-week study of the efficacy and safety of belimumab in patients of black African ancestry with systemic lupus erythematosus. *Arthritis Rheumatol* 2022;74:112–23.
 - 27 Phillippo DM. Multinma: Bayesian network meta-analysis of individual and aggregate data. R package version 0.5.0. Available: <https://zenodo.org/record/7033094#.Y8bhAHb7Q2w> [Accessed Oct 2022].
 - 28 Signorovitch JE, Sikirica V, Erder MH, *et al*. Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research. *Value Health* 2012;15:940–7.
 - 29 Phillippo DM, Ades T, Dias S, *et al*. Nice dsu technical support document 18: methods for population-adjusted indirect comparisons in submissions to NICE. Available: https://research-information.bris.ac.uk/ws/portalfiles/portal/94868463/Population_adjustment_TSD_FINAL.pdf [Accessed Oct 2022].
 - 30 Ballew N, Mian A, Levy RA, *et al*. Letter to the editor: indirect treatment comparison of anifrolumab efficacy versus belimumab in adults with systemic lupus erythematosus. *J Comp Eff Res* 2023;12.
 - 31 Bruce IN, Golam S, Steenkamp J, *et al*. Letter in reply: indirect treatment comparison of anifrolumab efficacy versus belimumab in adults with systemic lupus erythematosus. *J Comp Eff Res* 2023;12:e220192.
 - 32 Phillippo DM, Dias S, Ades AE, *et al*. Validating the assumptions of population adjustment: application of multilevel network meta-regression to a network of treatments for plaque psoriasis. *Med Decis Making* 2023;43:53–67.
 - 33 Murphy C-L, Yee C-S, Gordon C, *et al*. From BILAG to BILAG-based combined lupus assessment-30 years on. *Rheumatology (Oxford)* 2016;55:1357–63.
 - 34 Furie R, Wang L, Illei G, *et al*. Systemic Lupus Erythematosus (SLE) Responder Index response is associated with global benefit for patients with SLE. *Lupus* 2018;27:955–62.
 - 35 Thanou A, Chakravarty E, James JA, *et al*. Which outcome measures in SLE clinical trials best reflect medical judgment? *Lupus Sci Med* 2014;1:e000005.
 - 36 van Vollenhoven R, Askanase AD, Bomback AS, *et al*. Conceptual framework for defining disease modification in systemic lupus erythematosus: a call for formal criteria. *Lupus Sci Med* 2022;9:e000634.