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 Berstias,5, Alfred H J Kim,6 Antonis Fanouriakis,7, Roger A Levy,8 Deven Chauhan,9 Nick Ballew10

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 systematic 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. Treatment of SLE aims to minimise disease activity, decrease the incidence of disease flares and prevent organ damage. Conventional treatment options include antimalarials, glucocorticoids and immunosuppressive agents. Although such treatments can be initially successful, patients often require...
Belimumab, a human immunoglobulin monoclonal antibody, inhibits the biologic activity of B-lymphocyte stimulating protein. 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. 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). 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. Anifrolumab received approval based on evidence across three RCTs, two of which (MUSE 12 and TULIP-2 13) 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. 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. 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, however, the study did not meet the fundamental requirements for a robust population-adjusted analysis. Several studies have demonstrated that PAIC methods can perform poorly and yield inaccurate estimates under scenarios similar to that of the Bruce et al study. See Ballew et al (and the Discussion section) for further details on the limitations of the Bruce et al study. 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.

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. 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. 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)), 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. In the anifrolumab trials, SRI-4 incorporated SLEDAI-2K. Thus, in an effort to make as close as 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). Relatively, 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 al18 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. 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. 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. 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). 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
Table 1  Potential treatment EMs for SRI-4: characteristics that need to be balanced across trials

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

*Arrow thickness indicates the strength of the effect.
†Characteristic was adjusted for in sensitivity analyses.
‡Positivity in belimumab trials defined based on 30IU/mL threshold, while positivity in anifrolumab trials was defined based on 15IU/mL threshold. While it was possible to alter the definition in the belimumab trials to match the definition in the anifrolumab trials, a 15IU/mL threshold was not clinically meaningful for the belimumab trials.
§BMI of 27.6kg/m² for pooled TULIP trials (reported in Bruce et al⁷) and BMI of 25.4kg/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% CI) 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 al⁷ 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 24-point reduction in SLEDAI also suggested that belimumab and anifrolumab were generally comparable, with
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)
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. 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 approach, we obtained estimates in line with our primary analyses. These results are significantly different from those reported in Bruce et al. For example, whereas we obtained an SRI-4 OR of 1.11 with the MAIC, Bruce et al 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). Consequently, our comparison was made in the combined anifrolumab trial population and the Bruce et al 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. As reported in Bruce et al, 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 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, 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), 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. 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. However, the ML-NMR method we used does not suffer from the issues they allude to. 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. Thus, as we have demonstrated here, there is no need to remove eligible trials from the evidence base. Second, Bruce et al employed SRI-I values for the TULIP trials (24-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. One possible explanation for the discrepancy could be that Bruce et al employed results from a post hoc analysis of the TULIP-I SRI-I results. However, even if the revised post hoc definition for TULIP-I was used when pooling the trials, the OR would be 1.56. Third, Bruce et al 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. 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 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, 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. 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. 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...
compared 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
1Department of Rheumatology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
2Division of Rheumatology, Clinical Immunology and Allergy, University of Crete School of Medicine, Crete, Greece
3Division of Rheumatology, Department of Medicine, Washington University School of Medicine, Saint Louis, Missouri, USA
4First Department of Paedaeutic Internal Medicine, “Laikon” General Hospital, National Kapodistrian University of Athens Medical School, Athens, Greece
5Speciality Care, Global Medical Affairs, GSK, Collegeville, Pennsylvania, USA
6Division of Rheumatology, Clinical Immunology and Allergy, University of Crete School of Medicine, Crete, Greece
7Department of Rheumatology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
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 Fahrbach, 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 Thomson Reuters Scientific. MP has received grant/research support from AstraZeneca, Neupane, B., Lupus Science & Medicine.

REFERENCES