


Fragility of randomised controlled trials for systemic lupus erythematosus and lupus nephritis therapies

Gabriel Figueroa-Parra ¹, Michael S Putman,² Cynthia S Crowson,^{1,3} Alí Duarte-García ^{1,4}

To cite: Figueroa-Parra G, Putman MS, Crowson CS, *et al.* Fragility of randomised controlled trials for systemic lupus erythematosus and lupus nephritis therapies. *Lupus Science & Medicine* 2024;**11**:e001068. doi:10.1136/lupus-2023-001068

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/lupus-2023-001068>).

Preliminary results were presented at ACR Convergence 2023 (Figueroa-Parra G, Putman M, Duarte-García A. Fragility of randomized clinical trials of systemic lupus erythematosus and lupus nephritis therapies [abstract]. *Arthritis Rheumatol*. 2023; 75 (suppl 9). <https://acrabstracts.org/abstract/fragility-of-randomized-clinical-trials-of-systemic-lupus-erythematosus-and-lupus-nephritis-therapies/>).

Received 2 October 2023
Accepted 20 December 2023



© Author(s) (or their employer(s)) 2024. 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 Alí Duarte-García; duarte.ali@mayo.edu

ABSTRACT

Objective We aimed to evaluate the robustness of phase III randomised controlled trials (RCTs) for SLE and lupus nephritis (LN) using the fragility index (FI), the reverse FI (RFI) and the fragility quotient (FQ).

Methods We searched for phase III RCTs that included patients with active SLE or LN. Data on primary endpoints, total participants and the number of events for each arm were obtained. We calculated the FI score for RCTs with statistically significant results (number of patients required to change from event to non-event to make the study lose statistical significance), the RFI for RCTs without statistically significant results (number of patients required to change from non-event to event to make study gain statistical significance) and the FQ score for both (FI or RFI score divided by the sample size).

Results We evaluated 20 RCTs (16 SLE, four LN). The mean FI/RFI score was 13.6 (SD 6.6). There were nine RCTs with statistically significant results (seven SLE, two LN), and the mean FI score was 10.2 (SD 6.2). The lowest FI was for the ILLUMINATE-2 trial (FI=2), and the highest FI was for the BLISS-52 trial (FI=17).

Twelve studies had non-statistically significant results (10 SLE, two LN) with a mean RFI score of 15.6 (SD 6.1). The lowest RFI was for the ILLUMINATE-1 trial (RFI=4), and the highest RFI was for the TULIP-1 trial (RFI=27). The lowest FQ scores were found in the ILLUMINATE trials and the highest in the Rituximab trials (EXPLORER and LUNAR), meaning that the last ones were the most robust results after accounting for sample size.

Conclusions The evidence of therapies for patients with SLE and LN is derived mostly from fragile RCTs. Clinicians and trialists must be aware of the fragility of these RCTs for clinical decision-making and designing trials for novel therapeutics.

INTRODUCTION

SLE is a chronic autoimmune disease that primarily affects women and is characterised by heterogeneous clinical manifestations.¹ Its treatment aims to achieve remission, prevent organ damage, minimise drug side effects and improve quality of life.² The first US Food and Drug Administration (FDA)-approved therapy for patients with SLE was aspirin (1948), which was followed by glucocorticoids (1950s)

WHAT IS ALREADY KNOWN ON THIS TOPIC

- ⇒ Randomised controlled trials (RCTs) in SLE for novel therapeutics have frequently failed to meet the criteria for regulatory approval.
- ⇒ The fragility index aids in the interpretation of RCT findings and in identifying studies in which results could be overturned due to imprecision.

WHAT THIS STUDY ADDS

- ⇒ We found that the phase III, randomised, placebo-controlled trials in patients with SLE and lupus nephritis (LN) held an important degree of fragility.
- ⇒ Both positive and negative RCTs of SLE and LN therapeutics are not particularly robust.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

- ⇒ Clinicians should know that the evidence supporting several therapies in SLE and LN has been derived from fragile RCTs, and trialists should consider fragility among the challenges when designing clinical trials for patients with SLE and LN.

and the antimalarial drug hydroxychloroquine (1955).³ Despite important advances in therapy, which included the use of cyclophosphamide or mycophenolate mofetil for induction therapy of lupus nephritis (LN), the next approval took more than five decades.

Randomised controlled trials (RCTs) in SLE for novel therapeutics have frequently failed to meet criteria for regulatory approval. Even among recently approved drugs, trials have been mixed in their conclusions^{4 5} or have had small effect sizes for clinically relevant outcomes.^{6 7} Also, several RCTs of new therapies for patients with SLE or LN were terminated or were not published.^{8 9} In the last decade, the FDA approved three novel treatments for the management of SLE. First, in 2011, belimumab (a human monoclonal antibody that inhibits the B-cell activating factor) was approved for adult patients with non-renal SLE, which was followed in 2020 by its approval for LN. In 2021, voclosporin

(a calcineurin inhibitor) was approved for the treatment of LN, and anifrolumab (a human monoclonal antibody targeting type 1 interferon receptor) for adult patients with SLE.³ All four approvals were based on large double-blind RCTs that demonstrated superiority over matched placebo interventions. Although these therapies met the criteria for FDA approval,¹⁰ the robustness of the evidence supporting their use should be continuously assessed.

The fragility index (FI) is a recently described method for assessing the robustness of RCT findings, which calculates the minimum number of patients whose status would be required to change from an event to a non-event to make the study lose statistical significance.^{11,12} A small FI score indicates a more fragile and less statistically robust clinical trial; the fragility quotient (FQ) assesses fragility relative to a trial's sample size.¹² In this study, we aimed to assess the robustness of the published phase III RCTs for SLE and LN treatments using the FI, the reverse FI (RFI) and the FQ scores to complement the interpretation of these RCTs.

METHODS

To identify RCTs, we searched on ClinicalTrials.gov (up to 30 September 2023) for trials that included patients with SLE or LN. We included all phase III, randomised, placebo-controlled trials of patients with active SLE or LN. We excluded RCTs that were defined as phase II or IV RCTs as well as those whose status was recruiting or ongoing, completed but unpublished or terminated for reasons other than futility.

Data were obtained from the full-text publications on prespecified primary endpoints, total participants, participants in all arm doses for non-approved drugs and the arms of approved doses (in the USA), the number of events in the intervention and placebo groups, the number of patients who withdrew or discontinued and the reported p values. FI score was calculated using an online calculator (available at <https://clincalc.com/Stats/FragilityIndex.aspx>). The RFI was calculated for RCTs with non-statistically significant results by modifying the number of events in the intervention arm until a p value <0.05 was reached while keeping the total number of participants constant.¹³ The FQ score was calculated by dividing the FI (or RFI) score by the total sample size of the trial.¹² We used descriptive statistics to present the results. Analyses were performed using BlueSky Statistics software V.10.3 (BlueSky Statistics, Chicago, Illinois, USA).

RESULTS

We evaluated 20 RCTs, 16 in SLE and 4 in LN. The mean FI/RFI score of the 20 studies was 13.6 (SD 6.6). Additional characteristics of the included studies are shown in online supplemental table 1. There were nine RCTs with statistically significant results (seven in SLE, two in LN; table 1), which had a mean FI score of 10.2 (SD 6.2). The lowest FI was for the ILLUMINATE-2 trial (score:

2, tabalumab 120 mg every 4 weeks), and the highest FI was found in the BLISS-52 trial (score: 17, belimumab), meaning that would be required to have only two and 17 fewer responder patients, respectively, to lose significance.

Twelve studies showed non-statistically significant results (10 in SLE, two in LN; table 2) with a mean RFI score of 15.6 (SD 6.1). The lowest RFI (non-significant trials) was for the ILLUMINATE-1 trial (score: 4, tabalumab 120 mg every 4 weeks), and the highest RFI was for the TULIP-1 trial (score: 27, anifrolumab), meaning these trials would only need to have four and 27 more responders, respectively, to gain statistical significance.

Overall, the non-significant RCTs for SLE therapies seem to be more robust than those with statistically significant findings (figure 1). The RCTs for LN therapies showed a similar distribution.

The RCT that granted FDA approval to anifrolumab was the TULIP-2 trial, with an FI score of 11. Voclosporin trial (AURORA-1) had an FI score of 15. Regarding the belimumab studies, BLISS-52 had an FI score of 17, BLISS-76 had an FI score of 4, BLISS-SC had an FI score of 16 and BLISS-LN had an FI score of 3; all these trials had statistically significant results. The lowest FQ scores were found in the ILLUMINATE trials and the highest in the Rituximab trials (EXPLORER and LUNAR) and the early terminated LOTUS trial, suggesting that the last ones are the more robust results after accounting for sample size.

DISCUSSION

In this study of pivotal phase III, randomised, placebo-controlled trials of new treatments for SLE and LN, many trials had an important degree of fragility, even among RCTs of medications that obtained approval. These results may in part explain the perception that trials in SLE have frequently failed, which has been previously attributed to the complexity of the patients, the choice of the standard of care, the selection of the appropriate endpoints and the procedures needed for conditions with lower incidence and prevalence like SLE.^{14,15} Our results suggest that fragile trials—which require relatively few patients to switch groups to alter the conclusion of the study—may have also contributed.

Fragility was observed across all recently studied therapies. Among the approved drugs, the BLISS-LN trial was the most fragile, which would have been considered non-significant study (ie, a 'negative trial') if only three fewer patients in the treatment group had achieved the primary endpoint.¹⁶ This does not invalidate the approval of belimumab or suggest that physicians should not prescribe it. Rather, it highlights the limitations of the current methodological approach to conducting trials in SLE. These findings should be viewed in light of other concerning factors in BLISS-LN, which include the change in the original primary endpoint, the selected power of 80% to detect a difference in the endpoint or the underlying treatment that was left at the investigator's discretion (for

Table 1 Fragility evaluation of phase III, randomised, placebo-controlled trials with statistically significant results involving patients with active SLE and lupus nephritis (LN)

Study drug, trial	First author, year	Condition	Primary endpoints	Sample size	Events in intervention	Events in placebo	FI score	FQ score
Anifrolumab								
TULIP-2, NCT02446899 ⁵	Morand, 2020	Adults with active SLE	BICLA response at week 52	Anifrolumab 300 mg (n=180) Placebo (n=182)	86	57	11	0.030
Baricitinib								
SLE-BRAVE-I, NCT03616912 ¹⁹	Petri, 2023	Adults with active SLE	SRI-4 response at week 52	Baricitinib 2 mg (n=255)	126	116	15*	0.030
				Baricitinib 4 mg (n=252) Placebo (n=253)	142	4	0.008	
Belimumab								
BLISS-52, NCT00424476 ²⁸	Navarra, 2011	Adults with active SLE	SRI-4 response at week 52	Belimumab 10 mg/kg (n=290) Placebo (n=287)	167	125	17	0.029
BLISS-76, NCT00410384 ²⁹	Furie, 2011	Adults with active SLE	SRI-4 response at week 52	Belimumab 10 mg/kg (n=273) Placebo (n=275)	118	92	4	0.007
BLISS-SC, NCT01484496 ³⁰	Stohl, 2017	Adults with active SLE	SRI-4 response rate at week 52	Belimumab 200 mg subcutaneous (n=556) Placebo (n=280)	341	136	16	0.019
BEL113750, NCT01345253 ³¹	Zhang, 2018	Adults with active SLE	SRI-4 response rate at week 52	Belimumab 10 mg/kg (n=451) Placebo (n=226)	240	87	15	0.022
BLISS-LN, NCT01639339 ¹⁶	Furie, 2020	Adults with active LN	Primary efficacy renal response at week 104	Belimumab 10 mg/kg (n=223) Placebo (n=223)	96	72	3	0.007
Tabalumab								
ILLUMINATE-2, NCT01205438 ¹⁸	Merrill, 2016	Active SLE	SRI-5 response at week 52	Tabalumab 120 mg every 2 weeks (n=372)	143	104	15	0.020
				Tabalumab 120 mg every 4 weeks (n=376) Placebo (n=376)	131	2	0.002	
Voclosporin								
AURORA-1, NCT03021499 ⁷	Rovin, 2021	Adults with active LN	Complete renal response at week 52	Voclosporin (n=179) Placebo (n=178)	73	40	15	0.042

*Arm with non-significant results; results represent reverse FI score.
BICLA, BILAG (British Isles Lupus Assessment Group)-based composite lupus assessment; FI, fragility index; FQ, fragility quotient; SRI, SLE responder index.

both, the intervention and placebo arms). Studies have been designed under optimistic assumptions of efficacy, which is difficult given the heterogeneity of SLE. Given the precarity of these results, it seems reasonable to ask whether these results would be reproducible if the trial was run a second time or if a repeat trial would result in inconsistent findings, as has been the case for other recent therapies, including tabalumab,^{17 18} anifrolumab^{4 5} and baricitinib.^{19 20}

Conversely, the most robust result was from a negative trial (ie, not statistically significant), the TULIP-1 trial of anifrolumab. It would have required 27 more patients to have achieved an SLE responder index-4 response (the prespecified primary endpoint) for the trial to be considered positive. Despite the negative result from TULIP-1, the TULIP-2 trial was tailored to detect a significant BILAG (British Isles Lupus Assessment Group)-based composite lupus assessment response, based on secondary

Table 2 Fragility evaluation of non-statistically significant randomised clinical trials involving patients with active SLE and lupus nephritis (LN)

Study drug, trial	First author, year	Condition	Primary endpoints	Sample size	Events in intervention	Events in placebo	RFI score	FQ score
Abatacept								
NCT00430677 ³²	Furie, 2014	LN	Time to complete response	Abatacept 30/10 (n=99)	9	8	9	0.045
				Abatacept 10/10 (n=99) Placebo (n=100)	11		7	0.035
Anifrolumab								
TULIP-1, NCT02446912 ⁴	Furie, 2019	Adults with active SLE	SRI-4 response rate at week 52	Anifrolumab 300 mg (n=180) Placebo (n=184)	65	74	27	0.074
Atacicept								
APRIL-SLE, NCT00624338 ³³	Isenberg, 2015	Active SLE	Proportion of patients experiencing flare BILAG A or B at week 52	Atacicept 75 mg (n=159)	92	85	13	0.041
				Atacicept 150 mg (n=145) Placebo (n=157)	54 (discontinued)			
Baricitinib								
SLE-BRAVE-I, NCT03616912 ¹⁹	Petri, 2023	Adults with active SLE	SRI-4 response at week 52	Baricitinib 2 mg (n=255)	126	116	15	0.030
				Baricitinib 4 mg (n=252) Placebo (n=253)	142		4*	0.008
SLE-BRAVE-II, NCT03616964 ²⁰	Petri, 2023	Adults with active SLE	SRI-4 response at week 52	Baricitinib 2 mg (n=261)	120	116	22	0.043
				Baricitinib 4 mg (n=258) Placebo (n=256)	121		19	0.037
Blisibimod								
CHABLIS-SC1, NCT01395745 ³⁴	Merrill, 2018	Adults with active SLE	SRI-6 response at week 52	Blisibimod (n=245) Placebo (n=197)	115	83	13	0.029
Epratuzumab								
EMBODY1, NCT01262365 ³⁵	Clowse, 2017	Active SLE	BICLA response rate at week 48	Epratuzumab 600 mg every week (n=248)	93	85	14	0.028
				Epratuzumab 1200 mg every 2 weeks (n=244) Placebo (n=249)	97		9	0.018
EMBODY2, NCT01261793 ³⁵	Clowse, 2017	Active SLE	BICLA response rate at week 48	Epratuzumab 600 mg every week (n=264)	93	88	18	0.034
				Epratuzumab 1200 mg every 2 weeks (n=261) Placebo (n=263)	89		21	0.040
Rituximab								
EXPLORER, NCT00137969 ³⁶	Merrill, 2010	Active SLE	Major clinical response at week 52	Rituximab (n=169) Placebo (n=88)	50	25	20	0.078
LUNAR, NCT00282347 ³⁷	Rovin, 2012	Active LN	Renal response rate at week 52	Rituximab (n=72) Placebo (n=72)	19	22	16	0.111

Continued

Table 2 Continued

Study drug, trial	First author, year	Condition	Primary endpoints	Sample size	Events in intervention	Events in placebo	RFI score	FQ score
Tabalumab								
ILLUMINATE-1, NCT01196091 ¹⁷	Isenberg, 2016	Active SLE	SRI-5 response at week 52	Tabalumab 120 mg every 2 weeks (n=381)	121	111	17	0.022
				Tabalumab 120 mg every 4 weeks (n=378) Placebo (n=379)	133		4	0.005
Ustekinumab								
LOTUS†, NCT03517722 ³⁸	van Vollenhoven, 2022	Active SLE	SRI-4 response at week 52	Ustekinumab (n=173) Placebo (n=116)	76	51	22	0.076

*Arm with significant results; results represent FI score.
†Study terminated early due to futility.
BICLA, BILAG (British Isles Lupus Assessment Group)-based composite lupus assessment; FI, fragility index; FQ, fragility quotient; RFI, reverse FI; SRI, SLE responder index.

endpoints from TULIP-1. The FDA gave credence to the somewhat inexplicably different results from the TULIP-2 trial related primarily to the mucocutaneous and musculoskeletal involvement and granted approval. We observed two additional interventions that found mixed results in their trials. These were tabalumab (a human monoclonal antibody that binds B-cell activating factor) and baricitinib (a Janus kinase 1 and 2 inhibitor). The difference in the results from these studies might reflect the limitations of using endpoints based on global disease activity instruments, in contrast to system-based instruments for clinically heterogeneous conditions like SLE. Switching the main endpoints to a system-based approach might be beneficial, such as studying patients with specific manifestations (eg, arthritis or cutaneous disease) instead of having a global disease activity score that might result from several combinations of signs and symptoms. Perhaps these global disease activity scores could be kept as secondary outcomes. In other words, we should consider taking the same approach used for LN for

other SLE manifestations like was the case for the recent LILAC trial, which focused on patients with cutaneous and joint involvement.^{21 22} It would be interesting to test the fragility for future RCTs that used similar approaches by the two types of endpoints (global and system based).

Another issue for placebo-controlled trials in SLE throughout the years is the lack of truly standardised underlying therapy. Although it is currently recommended that all patients with SLE without contraindications must receive antimalarials,^{2 23 24} even in the most recent studies of anifrolumab and baricitinib, the proportion of patients on these medications did not go beyond 85% (the lowest 66–73% in TULIP-2). The use of other immunosuppressors, such as methotrexate, is also recommended (ie, for musculoskeletal and mucocutaneous manifestations, which are the most common across these RCTs) when antimalarials alone are insufficient, or high-dose glucocorticoids are required.^{2 23 24} However, these were not used in more than 20–25% of the patients from the most recent RCTs.^{4 5 19 20} This low use might be for different reasons, including intolerability or adverse events, patient or physician's decision, among other. If future clinical trials are system based as suggested above, it would be easier to agree on what is a standardised therapy for a specific disease manifestation.

One of the objectives of RCTs is to prove causality between the interventions being tested and the outcomes of interest. The most often used 'frequentist' approach relies on rejecting a null hypothesis (ie, there is not enough evidence of difference) based on a prespecified statistical significance threshold (typically $p < 0.05$). Requiring thresholds for statistical significance establishes an arbitrary dichotomy, whereby a trial with a p value of 0.049 is categorised similarly to one with a p value of 0.001. Stated within the framework of this paper, frequentist statistical approaches do not account for the fragility of a trial outcome. Alternative approaches to

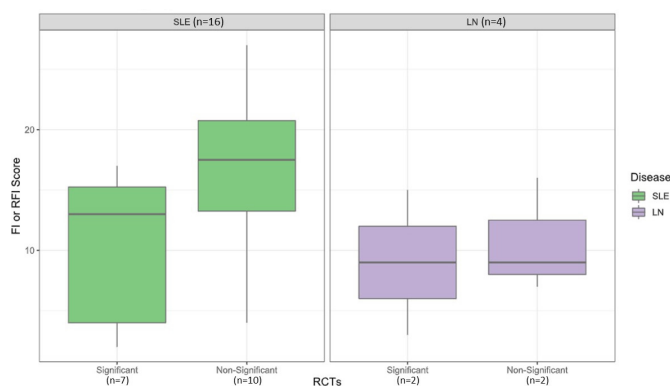


Figure 1 Fragility of randomised controlled trials (RCTs) for SLE and lupus nephritis (LN) therapies grouped by statistical significance of the primary endpoint. FI, fragility index; RFI, reverse FI.

trial design in SLE and LN may be considered, including Bayesian trial designs.²⁵ This approach presents other difficulties, including a lack of understanding of Bayesian methodology and interpretation among practising clinicians, ability for pharmaceutical sponsors to choose advantageous priors and the downstream issue that FDA approvals are necessarily dichotomous.

More simply, pivotal trials in SLE could be designed to be more robust. The easiest way to reduce the fragility of SLE trials would be through assuming a lower response rate among treatment groups. This would necessitate larger trials, but if it resulted in fewer negative trials for drugs that may have some efficacy (type II error), it would be a worthy trade-off. Another approach would be to select populations more likely to benefit. Finally, developing drugs that work better should be considered. Many of the aforementioned negative studies failed to surpass relatively low bars. In our efforts to understand why so many therapies have failed in SLE, we should consider the most obvious explanation; we do not understand SLE and our drugs do not work very well.

Alternative approaches for assessing available evidence exist. The American College of Rheumatology (ACR) guidelines have embraced the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach, for instance, which helps assess the quality of evidence and the strength of recommendations. The GRADE approach assigns the highest evidence grade to RCTs but can downgrade them if the studies have serious limitations, inconsistencies in the results, indirectness or imprecision concerns, along with the risk of reporting bias.²⁶ The FI is only one way of assessing imprecision within the GRADE domains²⁷ and must not be interpreted in isolation. The ACR has been undergoing efforts to update SLE and LN guidelines.

Our approach has limitations. First, there were relatively few RCTs available for each treatment. We did not evaluate secondary endpoints from the included trials because of the design-linked justification and due to the limited applicability of the FI/RFI to non-dichotomous endpoints. There are no validated approaches for interpreting the FI scores, which correlate with sample size (ie, larger studies likely to have larger FI scores), inversely correlate with p values (ie, small p values inevitably have larger FI scores) and may simply reflect good trial design (ie, small FI scores merely suggest that investigators correctly assessed the likely benefit of therapy and appropriately powered their study to reject the null hypothesis). FQ scores may account for this to some degree, but they are less intuitive than FI scores, and the interpretation of either score is ultimately subjective. Among our strengths is the novelty application of the FI approach in SLE and LN RCTs to aid clinicians and trialists in interpreting results from RCTs. We limited our evaluation to phase III trials for a more realistic picture of the currently approved treatments and those that matured enough in their pipeline to be compared with the standard of care.

These limitations notwithstanding, the results of this study suggest that the data informing approvals of novel SLE therapeutics may not be particularly robust. Comprehensive methodologies for evaluating the evidence supporting therapeutic interventions beyond the reductionist significance interpretation should be considered. In the interim, clinicians should know that the evidence supporting these therapies has mostly been derived from fragile RCTs, and trialists should consider fragility along with the rest of the challenges when designing clinical trials for patients with SLE and LN. Factors beyond the statistical fragility, like heterogeneity of clinical manifestations, characteristics of the selected endpoints and the lack of truly standardised therapy in the control arms (particularly in SLE RCTs), may have an impact on this analysis.

Author affiliations

¹Division of Rheumatology, Mayo Clinic, Rochester, Minnesota, USA

²Division of Rheumatology, Medical College of Wisconsin, Milwaukee, Wisconsin, USA

³Department of Quantitative Health Sciences, Mayo Clinic, Rochester, Minnesota, USA

⁴Robert D and Patricia E Kern Center for the Science of Health Care Delivery, Mayo Clinic, Rochester, Minnesota, USA

Twitter Gabriel Figueroa-Parra @DrGabrielFP, Michael S Putman @EBRheum, Cynthia S Crowson @CrowsonCindy and Ali Duarte-García @AliDuarteMD

Acknowledgements This manuscript is the product of the final project for the course 'Evidence-Based Medicine for Clinical Researchers' as part of the Postdoctoral Master's Degree in Clinical and Translational Science Program from the Center for Clinical and Translational Science and the Mayo Clinic Graduate School of Biomedical Sciences, where GFP is a scholar.

Contributors GF-P, MSP, CSC and AD-G contributed to the study conception and design. Material preparation and data collection were performed by GF-P. Analyses of data was performed by GF-P. Interpretation of results was made by GF-P, MSP, CSC and AD-G. The first draft of the manuscript was written by GF-P. GF-P and AD-G are responsible for the overall content of the study and act as guarantors. All authors critically revised for intellectual content and approved the final manuscript.

Funding This work was supported by the National Centre for Advancing Translational Sciences (grant number: UL1 TR002377).

Disclaimer Contents of the work are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

Competing interests AD-G has received unrelated grant funding from the Centers for Disease Control and Prevention, the Rheumatology Research Foundation Scientist Development Award and the Robert D and Patricia E Kern Center for the Science of Health Care Delivery.

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 This study is exempt from review or approval by the institutional review board due to the use of secondary data from published studies.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information.

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

Gabriel Figueroa-Parra <http://orcid.org/0000-0002-6077-8899>

Alí Duarte-García <http://orcid.org/0000-0003-1749-5719>

REFERENCES

- Fanouriakis A, Tziolos N, Bertsias G, et al. Update On the diagnosis and management of systemic lupus erythematosus. *Ann Rheum Dis* 2021;80:14–25.
- Fanouriakis A, Kostopoulou M, Alunno A, et al. Update of the EULAR recommendations for the management of systemic lupus erythematosus. *Ann Rheum Dis* 2019;78:736–45.
- FDA. *Lupus Therapies Continue to Evolve*. 2021. Available: <https://www.fda.gov/consumers/consumer-updates/lupus-therapies-continue-evolve>
- 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.
- Morand EF, Furie R, Tanaka Y, et al. Trial of Anifrolumab in active systemic lupus erythematosus. *N Engl J Med* 2020;382:211–21.
- Putman M. Anifrolumab in systemic lupus erythematosus. *N Engl J Med* 2020;382:1665–6.
- Rovin BH, Teng YKO, Ginzler EM, et al. Efficacy and safety of Voclosporin versus placebo for lupus nephritis (AURORA 1): a double-blind, randomised, Multicentre, placebo-controlled, phase 3 trial. *Lancet* 2021;397:2070–80.
- Pedersen C, Tai S, Valley E, et al. Unpublished clinical trials of common rheumatic diseases. *Rheumatology (Oxford)* 2023;62:3811–8.
- Gao Y, Wang Y, Li R, et al. n.d. Comprehensive analysis of clinical trials registration for lupus nephritis therapy on Clinicaltrials.Gov. *Front Med*;8.
- FDA. *Guidance for Industry Systemic Lupus Erythematosus - Developing Medical Products for Treatment* 2010. 2020. Available: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/systemic-lupus-erythematosus-developing-medical-products-treatment>
- Feinstein AR. "The unit fragility index: an additional appraisal of "statistical significance" for a contrast of two proportions". *J Clin Epidemiol* 1990;43:201–9.
- Tignanelli CJ, Napolitano LM. The fragility index in randomized clinical trials as a means of optimizing patient care. *JAMA Surg* 2019;154:74–9.
- Shi JL, Mojica ES, Moverman MA, et al. The reverse fragility index: interpreting the current literature on long-term survivorship of computer-Navigated versus conventional TKA: A systematic review and cross-sectional study of randomized controlled trials. *J Bone Joint Surg Am* 2023;105:157–63.
- Dall'Era M, Wofsy D. Clinical trial design in systemic lupus erythematosus. *Curr Opin Rheumatol* 2006;18:476–80.
- Touma Z, Gladman DD. Current and future therapies for SLE: obstacles and recommendations for the development of novel treatments. *Lupus Sci Med* 2017;4:e000239.
- Furie R, Rovin BH, Houssiau F, et al. Controlled trial of Belimumab in lupus nephritis. *N Engl J Med* 2020;383:1117–28.
- Isenberg DA, Petri M, Kalunian K, et al. Efficacy and safety of subcutaneous Tabalumab in patients with systemic lupus erythematosus: results from ILLUMINATE-1, a 52-week, phase III, Multicentre, randomised, double-blind, placebo-controlled study. *Ann Rheum Dis* 2016;75:323–31.
- Merrill JT, van Vollenhoven RF, Buyon JP, et al. Efficacy and safety of subcutaneous Tabalumab, a Monoclonal antibody to B-cell activating factor, in patients with systemic lupus erythematosus: results from ILLUMINATE-2, a 52-week, phase III, Multicentre, randomised, double-blind, placebo-controlled study. *Ann Rheum Dis* 2016;75:332–40.
- Morand EF, Vital EM, Petri M, et al. Baricitinib for systemic lupus erythematosus: a double-blind, randomised, placebo-controlled, phase 3 trial (SLE-BRAVE-I). *Lancet* 2023;401:1001–10.
- Petri M, Bruce IN, Dörner T, et al. Baricitinib for systemic lupus erythematosus: a double-blind, randomised, placebo-controlled, phase 3 trial (SLE-BRAVE-II). *Lancet* 2023;401:1011–9.
- Furie RA, van Vollenhoven RF, Kalunian K, et al. Trial of anti-Bdca2 antibody Litifilimab for systemic lupus erythematosus. *N Engl J Med* 2022;387:894–904.
- Werth VP, Furie RA, Romero-Diaz J, et al. Trial of anti-Bdca2 antibody Litifilimab for cutaneous lupus erythematosus. *N Engl J Med* 2022;387:321–31.
- Pons-Estel BA, Bonfa E, Soriano ER, et al. First Latin American clinical practice guidelines for the treatment of systemic lupus erythematosus: Latin American group for the study of lupus (GLADEL, Grupo Latino Americano de Estudio del Lupus)-Pan-American League of associations of rheumatology (PANLAR). *Ann Rheum Dis* 2018;77:1549–57.
- Mok CC, Hamijoyo L, Kasitanon N, et al. The Asia-Pacific League of associations for rheumatology consensus statements on the management of systemic lupus erythematosus. *The Lancet Rheumatology* 2021;3:e517–31.
- Berry DA. Bayesian clinical trials. *Nat Rev Drug Discov* 2006;5:27–36.
- Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. *BMJ* 2004;328:1490.
- Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines 6. rating the quality of evidence—Imprecision. *J Clin Epidemiol* 2011;64:1283–93.
- 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. *Ann Rheum Dis* 2018;77:355–63.
- Furie R, Nicholls K, Cheng T-T, et al. Efficacy and safety of Abatacept in lupus nephritis: a twelve-month, randomized, double-blind study. *Arthritis Rheumatol* 2014;66:379–89.
- Isenberg D, Gordon C, Licu D, et al. Efficacy and safety of Atacicept for prevention of flares in patients with moderate-to-severe systemic lupus erythematosus (SLE): 52-week data (APRIL-SLE randomised trial). *Ann Rheum Dis* 2015;74:2006–15.
- Merrill JT, Shanahan WR, Scheinberg M, et al. Phase III trial results with Blisibimod, a selective inhibitor of B-cell activating factor, in subjects with systemic lupus erythematosus (SLE): results from a randomised, double-blind, placebo-controlled trial. *Ann Rheum Dis* 2018;77:883–9.
- Clowse MEB, Wallace DJ, Furie RA, et al. Efficacy and safety of Epratuzumab in moderately to severely active systemic lupus erythematosus: results from two phase III randomized, double-blind, placebo-controlled trials. *Arthritis Rheumatol* 2017;69:362–75.
- 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 Rheum* 2010;62:222–33.
- Rovin BH, Furie R, Latinis K, et al. Efficacy and safety of Rituximab in patients with active proliferative lupus nephritis: the lupus nephritis assessment with Rituximab study. *Arthritis Rheum* 2012;64:1215–26.
- van Vollenhoven RF, Kalunian K, Dörner T, et al. Phase 3, Multicentre, randomised, placebo-controlled study evaluating the efficacy and safety of Ustekinumab in patients with systemic lupus erythematosus. *Ann Rheum Dis* 2022;81:1556–63.

Supplementary Table 1. Characteristics of included studies, total sample size, patients lost to follow-up and reported p-values.

Study drug, trial	Design	Arm	Sample size	Patients lost to follow-up/withdrawn	Reported p-value	FI	RFI
Abatacept, NCT00430677	Double-blind, parallel-group, multicenter (85 sites).	Abatacept, 30/10	298	25	0.806*	--	9
		Abatacept, 10/10	99	23	0.48*	--	7
		Placebo	100	22			
Anifrolumab, TULIP-1	Double-blind, parallel-group, multicenter (123 sites, 18 countries).	Anifrolumab, 300mg	457	36	0.412	--	27
		Placebo	180	38			
			184				
Anifrolumab, TULIP-2	Double-blind, parallel-group, multicenter (119 sites, 16 countries).	Anifrolumab, 300mg	365	27	0.001	11	--
		Placebo	180	52			
			182				
Atacicept, APRIL-SLE	Double-blind, parallel-group, multicenter (125 sites)	Atacicept, 75mg	461	45	0.543	--	13
		Placebo	159	43			
			157				
Baricitinib, SLE-BRAVE-I	Double-blind, parallel-group, multicenter (182 sites, 18 countries)	Baricitinib, 2mg	769	56	0.47	--	15
		Baricitinib, 4mg	258	55	0.016	4	--
		Placebo	254	68			
			257				
Baricitinib, SLE-BRAVE-II	Double-blind, parallel-group, multicenter (162 sites, 15 countries)	Baricitinib, 2mg	778	55	0.79	--	22
		Baricitinib, 4mg	262	62	0.71	--	19
		Placebo	259	49			
			257				
Belimumab, BLISS-52	Double-blind, parallel-group, multicenter (90 sites, 13 countries)	Belimumab, 10mg/kg	867	49	0.0006	17	--
		Placebo	290	61			
			287				

Study drug, trial	Design	Arm	Sample size	Patients lost to follow-up/withdrawn	Reported p-value	FI	RFI
Belimumab, BLISS-76	Double-blind, parallel-group, multicenter (136 sites, 19 countries)	Belimumab, 10mg/kg	819	82	0.017	4	--
		Placebo	273	89			
Belimumab, BLISS-SC	Double-blind, parallel-group, multicenter (177 sites, 30 countries)	Belimumab, 200mg SC	836	93	0.0006	16	--
		Placebo	556	66			
Belimumab, BEL113750	Double-blind, parallel-group, multicenter (49 sites, 3 countries)	Belimumab, 10mg/kg	707	82	<0.0001	15	--
		Placebo	451	56			
Belimumab, BLISS-LN	Double-blind, parallel-group, multicenter (107 sites, 21 countries)	Belimumab, 10mg/kg	448	76	0.03	3	--
		Placebo	223	91			
Blisibimod, CHABLIS-SC1	Double-blind, parallel-group, multicenter (15 countries)	Blisibimod	442	53	0.352	--	13
		Placebo	245	44			
Epratuzumab, EMBODY1	Double-blind, parallel-group, multicenter	Epratuzumab 600mg QW	793	94	0.442	--	14
		Epratuzumab 1200mg Q2W	248	81	0.175	--	9
		Placebo	244	90			
Epratuzumab, EMBODY2	Double-blind, parallel-group, multicenter	Epratuzumab 600mg QW	791	82	0.716	--	18
		Epratuzumab 1200mg Q2W	266	91	0.899	--	21
		Placebo	262	85			
Rituximab, EXPLORER	Double-blind, parallel-group, multicenter (55 sites)	Rituximab	257	49	0.886*	--	20
		Placebo	169	24			

Study drug, trial	Design	Arm	Sample size	Patients lost to follow-up/withdrawn	Reported p-value	FI	RFI																																																									
Rituximab, LUNAR	Double-blind, parallel-group, multicenter (52 sites)	Rituximab	144	7	0.55	--	16																																																									
		Placebo	72	18				Tabalumab, ILLUMINATE-1	Double-blind, parallel-group, multicenter	Tabalumab 120mg Q2W	1138	82	0.409	--	17	Tabalumab 120mg Q4W	381	87	0.052	--	4	Placebo	378	95				Tabalumab, ILLUMINATE-2	Double-blind, parallel-group, multicenter	Tabalumab 120mg Q2W	1124	77	0.002	15	--	Tabalumab 120mg Q4W	372	87	0.041*	2	--	Placebo	376	88				Ustekinumab, LOTUS	Double-blind, parallel-group, multicenter (140 sites, 20 countries)	Ustekinumab	516	155	1.00*	--	22	Placebo	308	103	Voclosporin, AURORA-1	Double-blind, parallel-group, multicenter (142 sites, 27 countries)	Voclosporin	357	15	<0.001
Tabalumab, ILLUMINATE-1	Double-blind, parallel-group, multicenter	Tabalumab 120mg Q2W	1138	82	0.409	--	17																																																									
		Tabalumab 120mg Q4W	381	87	0.052	--	4																																																									
		Placebo	378	95																																																												
Tabalumab, ILLUMINATE-2	Double-blind, parallel-group, multicenter	Tabalumab 120mg Q2W	1124	77	0.002	15	--																																																									
		Tabalumab 120mg Q4W	372	87	0.041*	2	--																																																									
		Placebo	376	88																																																												
Ustekinumab, LOTUS	Double-blind, parallel-group, multicenter (140 sites, 20 countries)	Ustekinumab	516	155	1.00*	--	22																																																									
		Placebo	308	103																																																												
Voclosporin, AURORA-1	Double-blind, parallel-group, multicenter (142 sites, 27 countries)	Voclosporin	357	15	<0.001	15	--																																																									
		Placebo	179	31																																																												

*Not reported in original article. Two-sided Fisher's exact test.