ARTHRITIS & RHEUMATOLOGY
Vol. 69, No. 4, April 2017, pp 785–790
DOI 10.1002/art.40013

BRIEF REPORT

Longitudinal Patterns of Response to Standard of Care Therapy for Systemic Lupus Erythematosus: Implications for Clinical Trial Design

Mimi Kim,[1] Joan Merrill,[2] Kenneth Kalunian,[3] Bevra Hahn,[4] Anita Roach,[5] and Peter Izmirly,[6] for the Lupus Foundation of America Collective Data Analysis Initiative Group

Objective. **To evaluate longitudinal patterns of response to standard of care for systemic lupus erythematosus (SLE) in clinical trials and to identify characteristics that differentiate nonresponders from persistent responders.**

Methods. **Data on 147 patients with moderately to severely active SLE without acute nephritis who were treated with placebo plus standard of care in two 52-week phase II/III trials were obtained from the Collective Data Analysis Initiative of the Lupus Foundation of America. Cross-sectional and longitudinal analyses of British Isles Lupus Assessment Group (BILAG)–based responses (improvement in all baseline A or B scores without new flare) were performed. Baseline characteristics that discriminated persistent responders from nonresponders were identified using logistic regression.**

Results. **Cross-sectional response rates decreased from 46% to 37% between 12 and 52 weeks. The overall rate of complete and sustained response, i.e., response at all visits, was only 14.3% (95% confidence interval 8.6–19.9%). Agreement between response status at 12 weeks and 36–52 weeks was low ($\kappa = 0.15$–0.29), and only 31% of initial 12-week responders maintained response at all subsequent visits. Baseline factors predictive of persistent response to standard of care included fewer organs with active disease, high C3 levels, and type of background therapy.**

Conclusion. **Use of sustained rather than landmark response may reduce high placebo response rates in SLE trials that continue aggressive standard of care. Further exploration to assess the power of this end point to improve discrimination between active and placebo arms is indicated. Lack of temporal stability in response highlights a potential weakness with shorter studies. Rates of response to standard of care are affected by the severity of the disease and the aggressiveness of background immunosuppressive treatments.**

Most clinical trials of new treatments for systemic lupus erythematosus (SLE) have failed to demonstrate robust discrimination between investigational agents and standard of care. Potential reasons include the heterogeneous manifestations of disease that can affect different organ systems, within-patient variability in disease activity during follow-up, weaknesses in primary end point and response definitions, and use of "add-on" trial designs in which the study drug or placebo is superimposed on aggressive background immunosuppressant regimens, resulting in high response rates in the control arms and little "room" to demonstrate efficacy of a potentially safer or more effective regimen. These issues have made it challenging to devise well-designed and sufficiently powered trials to detect effects of experimental treatments on clinically important outcomes. In addition, information about both the time of onset and the durability of response is critical for determining the appropriate length of patient follow-up and visit schedules. However, prior lupus trials have generally reported response rates based on cross-sectional analysis of data at landmark time points. Tracking landmark response rates over time is not the same as tracking durability of response, since the same people may not be responders at each visit. A better understanding of durability of response is also important for evaluating whether outcomes assessed

Supported by the Lupus Foundation of America.

[1]Mimi Kim, ScD: Albert Einstein College of Medicine, New York, New York; [2]Joan Merrill, MD: Oklahoma Medical Research Foundation, Oklahoma City; [3]Kenneth Kalunian, MD: University of California at San Diego, La Jolla, California; [4]Bevra Hahn, MD: University of California, Los Angeles; [5]Anita Roach, MS: Lupus Foundation of America, Washington, DC; [6]Peter Izmirly, MD: New York University School of Medicine, New York, New York.

Dr. Merrill has received consulting fees from Aspreva and Genentech (less than $10,000 each). Dr. Kalunian has received consulting fees from Genentech and Bristol-Myers Squibb (less than $10,000 each). Dr. Hahn has received a grant from Bristol-Myers Squibb.

Address correspondence to Mimi Kim, ScD, Albert Einstein College of Medicine, Belfer Building Room 1303, 1300 Morris Park Avenue, Bronx, NY 10461. E-mail: mimi.kim@einstein.yu.edu.

Submitted for publication June 20, 2016; accepted in revised form November 29, 2016.

earlier in follow-up are reliable interim end points or surrogate markers for longer term outcomes, and thus suitable for use in adaptive and other efficient trial designs.

A major goal of the Collective Data Analysis Initiative (CDAI) of the Lupus Foundation of America (LFA) is to improve future trials by learning from data from the placebo plus standard of care arms of completed phase II/III studies. In an earlier CDAI study, Kalunian et al (1) evaluated the effects of different background medications on British Isles Lupus Assessment Group (BILAG)–based response and flare rates (2). In the present study, we examine data from 52-week trials in the CDAI database to assess the longitudinal patterns of response to standard of care therapy, estimate the correlation in response status at different time points, and identify characteristics that differentiate nonresponders from persistent responders treated with placebo plus standard of care.

PATIENTS AND METHODS

The LFA established the CDAI, led by clinical investigators with experience in multicenter lupus trials, so that data from the placebo/standard of care arms of previous SLE clinical trials can be used to improve the design and conduct of future studies. Currently, the CDAI database includes deidentified data from 6 completed trials. For this analysis, we combined data from the standard of care/placebo arms of the two 52-week randomized phase II/III trials in patients with moderately to severely active lupus without acute nephritis. Trials that had shorter durations of follow-up were excluded. Patients with at least 1 BILAG A score or at least 2 BILAG B scores were included; one trial additionally allowed 1 BILAG B score at entry and required patients to have at least 1 BILAG A or B score in the musculoskeletal, mucocutaneous, and/or cardiovascular/respiratory organ systems. Immunosuppressive agents taken by most patients included azathioprine, mycophenolate mofetil, and methotrexate. Some patients also received antimalarial medications as background therapy, either alone or in combination with other treatments. Prednisone was initiated after screening in both trials at starting dosages of 30–60 mg/day and tapered 2–4 weeks later. Ethics committee approvals were obtained according to local regulations, and informed consent procedures were completed in accordance with the Helsinki Declaration version in effect at the time of each trial.

The classic BILAG index based on 8 index organ systems was used to define response, since it was the only disease activity measure that was common to both trials and is a sensitive measure that can capture partial or complete improvement in SLE. BILAG response was defined as an improvement of at least 1 grade compared to baseline in any organ with a BILAG A score (severely active) or B score (moderately active), without significant new disease, defined as at least 1 new A score or $\geq$2 new B scores. Response status was evaluated at weeks 12, 24, 36, 48, and 52. Baseline BILAG scores were also converted to numerical values, where A = 9, B = 3, C = 1, D = 0, and E = 0 (3). Scores were then summed across organ systems to yield a total baseline disease activity score for each patient.

Table 1. Baseline demographic and clinical characteristics of the patients with SLE (n = 147)*

Women, no. (%)	139 (95)
Race, no. (%)	
White	90 (61)
African American	28 (19)
Other	29 (20)
Age, years	39.4 ± 12.3
Weight, kg	74.8 ± 21.7
Height, cm	163.7 ± 8.1
BILAG score	13.1 ± 5.4
Baseline steroid dosage, mg/day	38.8 ± 17.4
Background medication at baseline, no. (%)†	
AZA	37 (26)
MMF	34 (23)
MTX	36 (25)
Other	38 (26)
Dipstick urinalysis for protein, no. (%)†	
0	130 (89)
1	9 (6)
2+	7 (5)
Organ system with a BILAG A or B score at baseline, no. (%)	
General	54 (37)
Mucocutaneous	95 (65)
Musculoskeletal	112 (76)
Cardiorespiratory	28 (19)
Neuropsychiatric	9 (6)
Vasculitis	16 (11)
Renal	1 (1)
Hematologic	24 (16)
Laboratory variables	
WBCs, $\times 10^3/\mu$l	7.4 ± 3.8
Lymphocytes, $\times 10^3/\mu$l	1.5 ± 1.1
Neutrophils, $\times 10^3/\mu$l	5.5 ± 3.3
Hemoglobin, gm/dl	12.2 ± 1.6
C3, gm/liter	0.96 ± 0.34
C4, gm/liter	0.16 ± 0.08
IgA, gm/liter	3.2 ± 1.6
IgG, gm/liter	13.7 ± 6.0
IgM, gm/liter	1.4 ± 1.1
IgG anticardiolipin, units/ml	12.2 ± 17.4
IgM anticardiolipin, units/ml	15.0 ± 19.8
Urinary protein:creatinine ratio	0.25 ± 0.60
Anti-dsDNA positive, no. (%)†	78 (54)

* Except where indicated otherwise, values are the mean ± SD. SLE = systemic lupus erythematosus; BILAG = British Isles Lupus Assessment Group; AZA = azathioprine; MMF = mycophenolate mofetil; MTX = methotrexate; WBCs = white blood cells; anti-dsDNA = anti–double-stranded DNA.
† Data were missing for some patients.

Pairwise agreement between response status at different time points was estimated by computing the kappa statistic. Logistic regression was performed to identify variables that discriminate persistent responders and never responders. Persistent responders were defined as patients who showed an initial response at 12 or 24 weeks and maintained that response over at least 3 subsequent visits; "never responders" were patients in whom a response was not achieved at any visit. Selection of variables to include in the final model was based on both statistical and clinical considerations. A variable for study/trial was

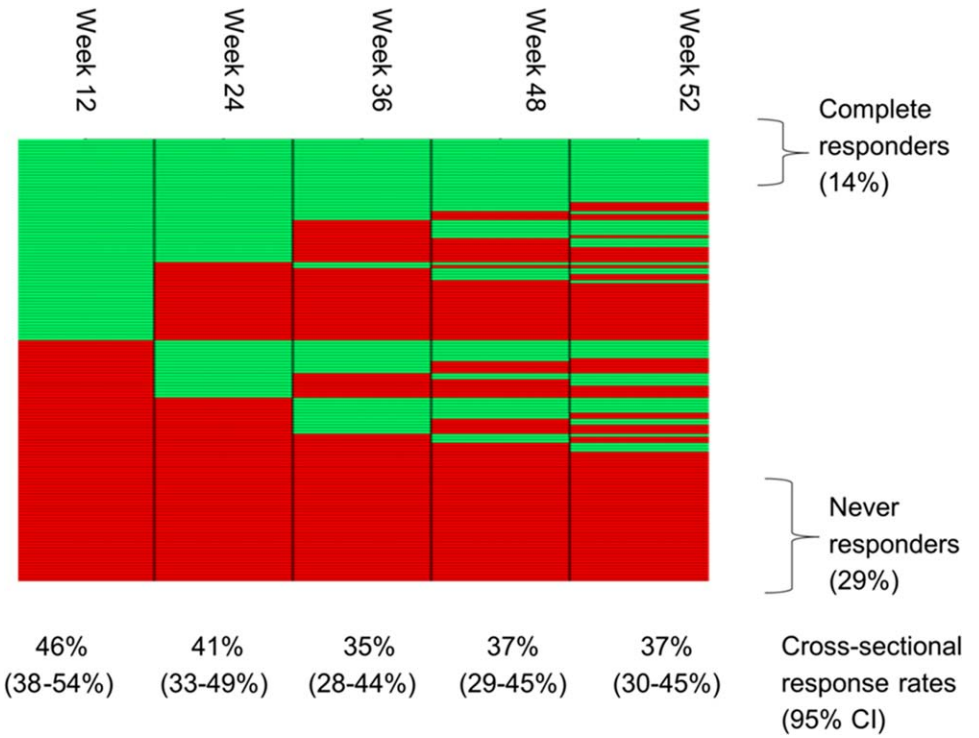

Figure 1. Temporal patterns in the response status of patients with systemic lupus erythematosus receiving standard of care. Green indicates response, and red indicates no response. 95% CI = 95% confidence interval.

included regardless of statistical significance to account for any systematic differences between the trials. Background immunosuppressant was also deemed to be clinically important given our prior findings (1). Given sample size constraints and nonlinear patterns in response rates across the total number of BILAG A and B scores, this variable was categorized as 1 A or 1 B score; 2 B scores; 2 A scores or 1 A and 1 B score; and ≥3 A or B scores. Missing data rates were 10.9%, 21.1%, 24.5%, 30.6%, and 29.3% at weeks 12, 24, 36, 48, and 52, respectively. Missing response status was imputed as nonresponse. Sensitivity analyses were also performed using only the observed data and by imputing missing data as response (best-case imputation). All analyses were performed in SAS, version 9.4. *P* values less than 0.05 (2-sided) were considered significant.

RESULTS

Baseline characteristics of pooled patients. The analysis included 147 patients (Table 1). Of these, 95% were women, 61% were white, and the mean ± SD age was 39.4 ± 12.3 years. The BILAG score at entry ranged from 3 to 33, and the mean ± SD score was 13.1 ± 5.4. As is typical in non-nephritis trials, disease activity was most commonly exhibited in the musculoskeletal and mucocutaneous domains, with 76% and 65% of subjects, respectively, having a BILAG score of A or B in those organs. Fifty-six percent had at least 1 BILAG score of A in any organ at baseline. The mean ± SD steroid dosage at baseline was 38.8 ± 17.4 mg/day of prednisone or equivalent. Nearly equal proportions of patients were taking azathioprine (26%), mycophenolate mofetil (23%), methotrexate (25%), and other treatments (26%) as background medications. Among the 38 patients in the "other" treatment category, 25 (66%) received antimalarials as primary background therapy; 73% of patients overall (107 of 147) were taking antimalarials with or without other medications.

Patterns of response status during follow-up. The cross-sectional response rates at each visit are shown at the bottom of Figure 1. The rate of response to standard of care was 46% (95% confidence interval [95% CI] 38–54%) at week 12 and steadily decreased over time to 37% (95% CI 30–45%) at week 52. Also shown are the longitudinal patterns of response status, where each row of the heatmap displays the response profile for 1 subject. Only 14% (21 of 147; 95% CI 8.6–19.9%) were "complete" responders who achieved and maintained a response at all 5 visits; 25% (37 of 147; 95% CI 18.2–32.2%) were persistent responders, defined as having an initial response at 12 or 24 weeks with maintenance of response at 3 or more subsequent visits. Both the complete and persistent response rates were much lower than any of the cross-sectional response rates. Furthermore, 29% (43 of 147) did not respond at any visit, and an additional 21% (31 of 147) responded at only a single visit.

Table 2. Logistic regression analysis of predictors of persistent response to standard of care for SLE*

Predictor variable	OR (95% CI)	*P*
Trial: A vs. B	1.23 (0.15–9.93)	0.85
Background medication		
AZA	0.96 (0.089–10.27)	0.97
MMF	0.19 (0.014–2.65)	0.22
MTX	0.05 (0.005– 0.57)	0.02
Other	1	–
C3 (per SD increase)	2.68 (1.27–5.67)	0.01
No. of BILAG A and B scores		
1: A or B	25.31 (3.34–192.0)	0.002
2: A/A or A/B	7.46 (1.19–46.95)	0.03
2: B/B	2.14 (0.41–11.17)	0.37
≥3: As or Bs	1	–

* The C statistic for the model was 0.87. SLE = systemic lupus erythematosus; OR = odds ratio; 95% CI = 95% confidence interval; AZA = azathioprine; MMF = mycophenolate mofetil; MTX = methotrexate; BILAG = British Isles Lupus Assessment Group.

The majority of the patients who had a response at any time during follow-up ("ever responders") first responded at 12 weeks (67 of 104 [64%]). However, only a third of the initial 12-week responders (21 of 67 [31%]) had a sustained response at each follow-up visit; nearly an equal proportion responded only at 12 weeks (19 of 67 [28%]). Eighteen percent of ever responders first responded at 24 weeks; among these, only about a third maintained the response through week 52, and 21% demonstrated a response only at 24 weeks. Less than 12% (12 of 104) of ever responders first responded at 36 weeks, and only 6% (6 of 104) first responded at 48 weeks or later. Early maintenance of response was an indicator of later response. Among 27 subjects who consistently responded from 12 to 36 weeks, ~80% (21 of 27) maintained that response through 52 weeks.

The degree of agreement was low to moderate between response status at week 12 and subsequent weeks, and the kappa statistics consistently declined over time ($\kappa = 0.38$ for agreement between response status at week 12 and week 24, and $\kappa = 0.15$–0.29 for agreement between response status at week 12 and weeks 36–52). Kappa values between response status at week 24 and later weeks were higher but did not exceed 0.5. Agreement between response status at week 36 and subsequent weeks ranged from 0.49 to 0.59 and was slightly higher between response status at 48 weeks and response status at 52 weeks ($\kappa = 0.66$). Corresponding kappa values did not differ by more than 0.10 in sensitivity analyses using the observed data and under best-case imputation.

Predictors of persistent response versus nonresponse. Logistic regression analysis was performed to identify independent variables measured at baseline that discriminate the 37 patients who were persistent responders and the 43 who did not respond at any visit, i.e., nonresponders (Table 2). Not surprisingly, features consistent with less active disease at screening were associated with persistent response, including fewer organs with active disease and higher C3 levels. However, among those with 2 organs with active disease, the odds of persistent response was greater among those with at least 1 BILAG A score compared to those with 2 BILAG B scores, perhaps reflecting the effects of more aggressive treatment intervention in patients with more severe manifestations. Methotrexate as background therapy was significantly associated with nonresponse. The C statistic for the model was 0.87, indicating high ability of the model to discriminate between persistent responders and nonresponders. Sensitivity analyses using available data only and under best-case imputation yielded similar findings, with fewer organs with active disease and high C3 levels associated with persistent response (data not shown).

DISCUSSION

Most clinical trials of SLE published to date have followed up patients for 52 weeks or more, but an increasing number of early-phase trials are terminating at 12 or 24 weeks (4). Our findings underscore the risk of using results from shorter trials to design future 52-week studies that include a standard of care arm, since only a third of patients who initially responded to standard of care at 12 and 24 weeks maintained that response at all subsequent visits. Moreover, 36% of patients in whom a response was achieved at any visit had their initial response after 12 weeks, and so would have been missed in a 12-week trial. Although shorter trials may be necessary in earlier phases of therapeutic development, it is important to design later phases with appropriate end points that can better predict success and long-term sustainability of response. Response to standard of care at 36 weeks correlated well with response status at 52 weeks, and 80% of patients who maintained a response between 12 and 36 weeks became complete responders through week 52. Therefore, 36 or 48 weeks of follow-up may be sufficient in future trials, although onset of efficacy and durability may differ in patients exposed to experimental therapies.

The observed lack of stability in response status over time implies that the duration of response should also be considered in end points when evaluating the effects of treatments. If response is defined as both achieving a response at a landmark time (e.g., 52 weeks) based on improvement from baseline and the absence of

interim flares, a patient who did not experience a flare but who also did not show any improvement until week 52 would still be considered a responder even if response was achieved only at that visit. A more stringent definition of response that incorporates response duration would minimize the false-positive detection of transient improvements and hopefully reduce the high standard of care response rates that may have contributed to the failure of past lupus trials.

One advantage of a lower placebo response rate is that the required sample size can be smaller to detect the same absolute difference in response rates between treatment groups. For example, 388 patients per arm are needed for 80% power to detect an increase in the response rate from 40% in the placebo arm to 50% in the experimental arm, whereas 250 patients per arm are needed to detect an increase from 15% to 25%. Several studies found that standard of care response rates were reduced and treatment differences magnified when the response criteria were based on higher thresholds for improvement in disease activity, e.g., ≥8-point decrease in Safety of Estrogens in Lupus Erythematosus National Assessment version of the SLE Disease Activity Index (SLE Responder Index 8 [SRI-8]) compared to SRI-5 (5,6). Whether a more stringent response definition that also includes duration of response will yield more powerful studies and better differentiation between effective medications and standard of care needs to be explored with additional data from both the experimental and placebo arms of past trials. When the response rates were compared across background medications in our CDAI data, the difference between the treatments with the lowest response rate (methotrexate) and the highest response rate (other medications) at 52 weeks increased from 19% with cross-sectional response ($P = 0.10$) to 30% with sustained response ($P = 0.005$), indicating better discrimination between treatments with sustained response.

In addition to considering end points based on higher thresholds for response, one should also target for inclusion in SLE trials patients who are less likely to respond to the rescue therapy that is allowed to be added to background medications. Our finding that lower placebo rate responses are generally associated with factors indicating greater disease is not a surprise and confirms the results of previous studies (3,7–9). Furthermore, our results are consistent with subgroup analyses of prior trials suggesting that higher disease activity at baseline and low complement levels improve discrimination between active treatments and placebo (6,10–13). In populations with moderate disease, decreasing or minimizing background therapies is proving to be a promising approach for identifying effective new therapies (14). Unfortunately, this alternative is not acceptable for patients at high risk of severe or organ-threatening disease.

Prior trials in lupus were generally conducted as conventional randomized 2-arm superiority studies. Alternative trial designs such as adaptive trials should be considered to improve trial efficiency. However, adaptive trials that use interim results to modify an ongoing study rely on the availability of end points that can be assessed quickly. The lack of agreement we observed in response status across follow-up visits implies that cross-sectional 12-week or 24-week status may not be a reliable surrogate marker for longer-term clinical outcomes and that more predictive interim end points need to be identified to conduct adaptive trials in lupus.

In this study, we used a BILAG-based response because SLEDAI and physician's global assessment were not measured in all trials. Results could differ for alternative response definitions, although the directionality of results tends to be similar with other end points. In addition, our findings may not be generalizable to patient populations who have different baseline characteristics or are exposed to other background medications. We hope to confirm these results with new data from the growing CDAI database and better elucidate the response patterns and predictors of response to standard of care.

ACKNOWLEDGMENT

We thank Dr. Kith Pradhan for assistance with generating Figure 1.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Kim had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Kim, Merrill, Kalunian, Hahn, Izmirly.
Acquisition of data. Kim, Roach.
Analysis and interpretation of data. Kim, Merrill, Kalunian, Hahn, Izmirly.

REFERENCES

1. Kalunian KC, Kim M, Xie X, Baskaran A, Daly RP, Merrill JT. Impact of standard of care treatments and disease variables on outcomes in systemic lupus erythematosus trials: analysis from the Lupus Foundation of America Collective Data Analysis Initiative. Eur J Rheumatol 2016;3:13–9.
2. Hay EM, Bacon PA, Gordon C, Isenberg DA, Maddison P, Snaith ML, et al. The BILAG index: a reliable and valid instrument for measuring clinical disease activity in systemic lupus erythematosus. Q J Med 1993;86:447–58.
3. Merrill JT, Neuwelt CM, Wallace DJ, Shanahan JC, Latinis KM, Oates JC, et al. Efficacy and safety of rituximab in moderately-to-severely active systemic lupus erythematosus: the randomized, double-blind, phase II/III systemic lupus erythematosus evaluation of rituximab trial. Arthritis Rheum 2010;62:222–33.

4. Kalunian KC, Merrill JT, Maciuca R, McBride JM, Townsend MJ, Wei X, et al. A phase II study of the efficacy and safety of rontalizumab (rhuMAb interferon-α) in patients with systemic lupus erythematosus (ROSE). Ann Rheum Dis 2016;75: 196–202.
5. Furie RA, Leon G, Thomas M, Petri MA, Chu AD, Hislop C, et al. A phase 2, randomised, placebo-controlled clinical trial of blisibimod, an inhibitor of B cell activating factor, in patients with moderate-to-severe systemic lupus erythematosus, the PEARL-SC study. Ann Rheum Dis 2015;74:1667–75.
6. Khamashta M, Merrill JT, Werth VP, Furie R, Kalunian K, Illei GG, et al. Sifalimumab, an anti-interferon-α monoclonal antibody, in moderate to severe systemic lupus erythematosus: a randomised, double-blind, placebo-controlled study. Ann Rheum Dis 2016;75:1909–16.
7. Mysler EF, Spindler AJ, Guzman R, Bijl M, Jayne D, Furie RA, et al. Efficacy and safety of ocrelizumab in active proliferative lupus nephritis: results from a randomized, double-blind, phase III study. Arthritis Rheum 2013;65:2368–79.
8. Petri MA, van Vollenhoven RF, Buyon J, Levy RA, Navarra SV, Cervera R, et al. Baseline predictors of systemic lupus erythematosus flares: data from the combined placebo groups in the phase III belimumab trials. Arthritis Rheum 2013;65:2143–53.
9. Wallace DJ, Popa S, Spindler AJ, Eimon A, González-Rivera T, Utset TO, et al. Improvement of disease activity and reduction of severe flares following subcutaneous administration of an IL-6 monoclonal antibody (mAb) in subjects with active generalized systemic lupus erythematosus (SLE) [abstract]. Arthritis Rheumatol 2014;66:3531.
10. Furie R, Merrill J, Werth V, Khamashta M, Kalunian K, Brohawn P, et al. Anifrolumab, an anti-interferon α receptor monoclonal antibody, in moderate to severe systemic lupus erythematosus (SLE) [abstract]. Arthritis Rheumatol 2015;67 Suppl 10. URL: http://acrabstracts.org/abstract/anifrolumab-an-anti-interferon-alpha-receptor-monoclonal-antibody-in-moderate-to-severe-systemic-lupus-erythematosus-sle/.
11. Isenberg D, Gordon C, Licu D, Copt S, Rossi CP, Wofsy D. 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.
12. Merrill JT, van Vollenhoven RF, Buyon JP, Furie RA, Stohl W, Morgan-Cox M, 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.
13. Van Vollenhoven RF, Petri MA, Cervera R, Roth DA, Ji BN, Kleoudis CS, et al. Belimumab in the treatment of systemic lupus erythematosus: high disease activity predictors of response. Ann Rheum Dis 2012;71:1343–9.
14. Merrill JT, Immermann FW, Zhou T, O'Toole M, Whitley M, Hill AA, et al. Topline results of the Biomarkers of Lupus Disease (BOLD) study: clinical and mechanistic perplexities of lupus treatment trials can be mitigated by eliminating background immune suppressants [abstract]. Arthritis Rheum 2013;65 Suppl 10:S771–2.