UCSF
UC San Francisco Previously Published Works

Title

Comparison of Alternative Primary Outcome Measures for Use in Lupus Nephritis Clinical Trials

Permalink

https://escholarship.org/uc/item/90p19946

Journal

Arthritis & Rheumatism, 65(6)

ISSN

0004-3591

Authors

Wofsy, David
Hillson, Jan L
Diamond, Betty

Publication Date

2013-06-01

DOI

10.1002/art.37940

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

ARTHRITIS & RHEUMATISM
Vol. 65, No. 6, June 2013, pp 1586–1591
DOI 10.1002/art.37940
© 2013, American College of Rheumatology

Comparison of Alternative Primary Outcome Measures for Use in Lupus Nephritis Clinical Trials

David Wofsy,[1] Jan L. Hillson,[2] and Betty Diamond[3]

Objective. **Clinical trials of therapies for lupus nephritis have used many different primary outcome measures, ranging from complete response to time to end-stage renal disease. The objective of this study was to compare several possible outcome measures, using data from a large, multicenter trial of abatacept in lupus nephritis, to gain insight into which outcome measure, if any, was best able to discern differences among treatment groups.**

Methods. **Study patients received either abatacept or placebo, on a background of mycophenolate mofetil and glucocorticoids. Using data from this trial, the following primary outcome measures at 24 and 52 weeks were compared: complete response rate, major clinical response rate, total response rate (complete plus partial response), improvement in proteinuria, improvement in estimated glomerular filtration rate, and frequency of treatment failure. Time to complete response was also evaluated.**

Results. **Complete response rate, major clinical response rate, and time to complete response were the measures that best discriminated between the abatacept groups and placebo, and the sensitivities of these 3 measures were comparable. For these measures, sample sizes of 50 patients would have been sufficient to demonstrate a statistically significant difference between treatment and control at 52 weeks. Each of the other measures also discriminated between treatment and control, but much larger group sizes would have been required to determine statistical significance.**

Conclusion. **The choice of primary outcome measure can substantially influence the ability to detect therapeutic benefit in lupus nephritis trials. This study suggests that complete response rate, major clinical response rate at 52 weeks, and time to complete response may be the most sensitive outcome measures for detecting differences among therapeutic regimens.**

We recently examined the data from a large, multicenter, randomized, double-blind, placebo-controlled trial of abatacept in patients with lupus nephritis to evaluate the performance of several definitions of complete response (CR) (1). Our analysis highlighted that there was great variation in the ability of the different definitions of CR to discriminate between treatment groups. Our work provided a compelling demonstration that, in some instances, the choice of outcome measure, rather than the actual data, may determine whether an experimental agent is perceived as effective or ineffective. However, because the analysis was limited to various definitions of CR, it did not address whether CR should be the preferred outcome measure when testing novel therapies for patients with active lupus nephritis, or whether other outcome measures could discriminate more successfully among treatment groups.

Previous lupus nephritis trials have used a wide range of primary outcome measures. The landmark trial of pulse cyclophosphamide conducted at the National Institutes of Health focused on the frequency of end-stage renal disease (2). The Euro-Lupus Nephritis Trial

ClinicalTrials.gov identifier: NCT00430677. This is protocol IM101075 from Bristol-Myers Squibb.

Supported by the Rosalind Russell Medical Research Center for Arthritis at the University of California, San Francisco.

[1]David Wofsy, MD: University of California, San Francisco; [2]Jan L. Hillson, MD: Bristol-Myers Squibb, Princeton, New Jersey; [3]Betty Diamond, MD: Feinstein Institute for Medical Research, Manhasset, New York.

Dr. Wofsy has received consulting fees from Genentech, Human Genome Sciences, and Vifor Pharma (less than $10,000 each) and from Bristol-Myers Squibb (more than $10,000). Dr. Hillson owns stock or stock options in Bristol-Myers Squibb. Dr. Diamond has received consulting fees and/or honoraria from Genentech, Eli Lilly, Merck-Serono, Bristol-Myers Squibb, Pfizer, and Astellas Pharma (less than $10,000 each). Drs. Wofsy and Diamond were not compensated for this investigator-initiated study.

Address correspondence to David Wofsy, MD, Arthritis/Immunology Unit (111R), San Francisco VA Medical Center, 4150 Clement Street, San Francisco, CA 94121. E-mail: wofsyd@medsch.ucsf.edu.

Submitted for publication October 9, 2012; accepted in revised form March 12, 2013.

Table 1. Definitions of potential outcome measures*

Outcome measure and definition
Complete response
UPCR <0.5.
For patients with abnormal serum creatinine levels at baseline, return to normal levels.
For patients with normal serum creatinine levels at baseline, final value within 15% of baseline value.
Inactive urine sediment (defined as <5 RBCs, <5 WBCs, no cellular casts).
Partial response
For patients with UPCR of >3 at baseline, reduction in UPCR to <3.
For patients with UPCR of ≤3 at baseline, reduction in UPCR of at least 50%, to final UPCR of <1.
Serum creatinine level within 15% of baseline level.
Major clinical response
Either of the following:
Patients meeting the complete response criteria (as defined above); or
Patients who had nephrotic-range proteinuria (UPCR >3) at baseline and who achieved a UPCR <1 at end point, and who met all of the other criteria for a complete response.
Time to complete response
Among patients who met the complete response criteria at week 52, the first time point at which the patient met the complete response criteria.
Improvement in proteinuria
UPCR reduced by ≥75% compared with baseline value.
Improvement in eGFR
eGFR increased by ≥25% compared with baseline value.
Treatment failure
Any of the following:
Persistent nephrotic-range levels of proteinuria (UPCR >3).
Failure of UPCR to improve by >25%.
eGFR abnormally low and reduced by >25% relative to baseline value.
Failure to taper glucocorticoid dosage to ≤10 mg/day.
Withdrawal due to worsening nephritis, infection, or drug toxicity.

* Except for treatment failure, all other outcome measures required that the glucocorticoid dosage be tapered to ≤10 mg/day of prednisone (or equivalent). UPCR = urine protein:creatinine ratio; RBCs = red blood cells; WBCs = white blood cells; eGFR = estimated glomerular filtration rate.

of mini-pulse cyclophosphamide therapy also relied on treatment failure as the primary outcome measure, although the definition of treatment failure used was broader than just end-stage renal disease (3). More recently, trials of new agents have shifted the paradigm from frequency of treatment failure to measures of treatment success, permitting an earlier assessment of response (1,4,5). Even in these studies, however, the choice of primary outcome has differed, with some studies relying on CR rates, some on total response rates (CR plus partial response [PR]), and some on time to response. Furthermore, whereas some studies examined the primary end point at 6 months, others examined the primary end point at 12 months.

This study was undertaken to assess the relative discriminatory capability of a number of potential outcome measures for use in lupus nephritis clinical trials. To accomplish this goal, we applied a diverse set of outcome measures to the same clinical data (1).

PATIENTS AND METHODS

Study design. The trial was a 12-month, phase II/III, multicenter, randomized, double-blind, controlled trial of abatacept versus placebo, on a background of standard-of-care treatment with mycophenolate mofetil (MMF) and glucocorticoids, in patients with active lupus nephritis. The study design has been described in detail previously (1). Briefly, 300 patients who met, either sequentially or coincidentally, 4 of the 11

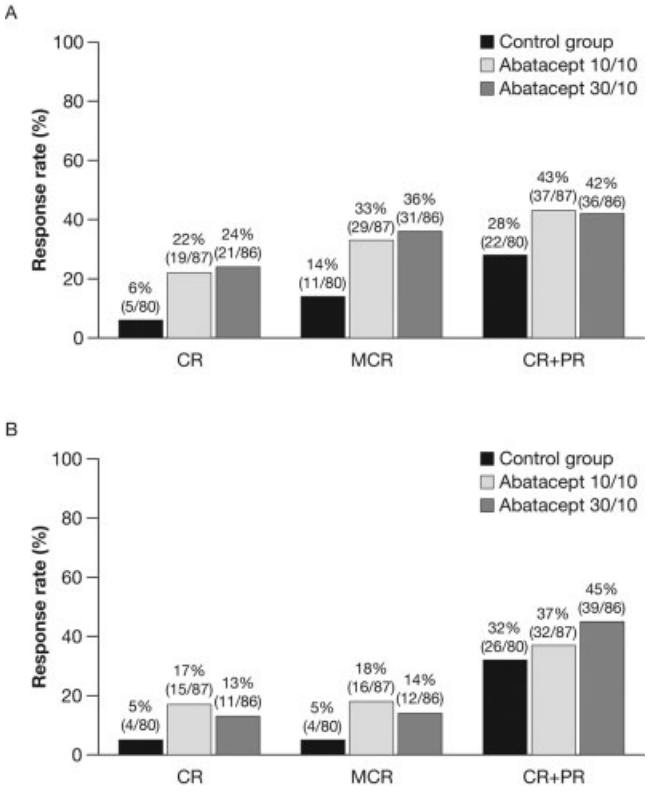

Figure 1. Complete response (CR) rates, major clinical response (MCR) rates, and total response (CR plus partial response [PR]) rates at week 52 **(A)** and week 24 **(B)** in each abatacept treatment group compared with the standard-of-care placebo control group. For the treatment groups, one group received abatacept infusions on days 1, 15, and 29, and every 28 days thereafter for 12 months at a fixed, weight-tiered dose of ~10 mg/kg (abatacept 10/10), while the other group received a higher dose of abatacept for the first 5 infusions (30 mg/kg), followed by a fixed, weight-tiered dose of ~10 mg/kg every 28 days thereafter (abatacept 30/10).

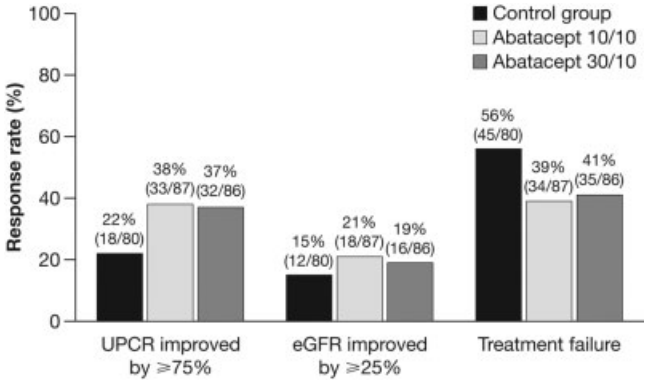

Figure 2. Frequency of improved urine protein:creatinine ratio (UPCR) (reduction in UPCR of ≥75%), improved estimated glomerular filtration rate (eGFR) (increase in eGFR of ≥25%), and treatment failure at week 52 in each abatacept treatment group (as described in Figure 1) compared with the standard-of-care placebo control group.

classification criteria for systemic lupus erythematosus, as defined by the American College of Rheumatology (6), and who had active, biopsy-proven class III or class IV lupus nephritis were randomly assigned 1:1:1 to 1 of 3 groups; 2 patients did not receive study medication. One group received placebo infusions on days 1, 15, and 29, and every 28 days thereafter for 12 months. A second group received abatacept infusions at a fixed, weight-tiered dose approximating 10 mg/kg according to the same schedule (abatacept 10/10). The third group received a higher dose of abatacept for the first 5 infusions (30 mg/kg), followed by a fixed, weight-tiered dose of ~10 mg/kg every 28 days (abatacept 30/10). All patients also received MMF and glucocorticoids throughout the trial. Oral glucocorticoids were initiated at a dosage equivalent to 30–60 mg/day of prednisone or prednisolone. A tapering regimen was recommended that was designed to reach a dosage of 10 mg/day by week 12 of treatment, but adherence to this regimen was not required if the site investigator deemed it not in the patient's best interest.

Comparison of outcome measures. Table 1 describes the outcome measures that were applied to the data set. The CR and PR criteria were obtained from the Lupus Nephritis Assessment of Rituximab (LUNAR) trial of rituximab for lupus nephritis (5). Those criteria were chosen because the LUNAR definition of CR discriminated most clearly between treatment and control in our prior comparison of various CR definitions (1). We also examined a closely related criterion, termed major clinical response (MCR). This possible alternative to CR was included to acknowledge that some of the "partial responses" observed in patients with very severe nephritis at baseline appeared to be at least as dramatic as some of the CRs in patients with mild disease at baseline.

Previous studies have shown that a 75% reduction in proteinuria correlates well with a favorable long-term outcome in patients with lupus nephritis (7). Therefore, we used a reduction of at least 75% as the threshold for evaluating improvement in proteinuria. With regard to renal function, there are no data in the literature to provide a clear rationale for choosing a particular threshold for improvement, so we defined improvement in renal function somewhat arbitrarily as an increase of at least 25% in the estimated glomerular filtration rate (eGFR).

Finally, we also examined the time to CR and the frequency of treatment failure, as defined in Table 1. For the purpose of these analyses, we based our definition of time to CR on the principle that the concept of time to CR is only meaningful for patients who maintained a status of CR through week 52 and not for patients in whom CR was evanescent. Therefore, the analysis of time to CR reflects only those patients who met the criteria for CR at week 52.

As noted above, the CR and PR criteria used in our analyses were drawn from the LUNAR trial. This trial required a urine protein:creatinine ratio (UPCR) of ≥1.0 for patient inclusion. Therefore, we limited our analyses to the subset of patients who met this entry criterion, which resulted in the following group sizes: control (n = 80), abatacept 10/10 (n = 87), and abatacept 30/10 (n = 86).

Statistical analysis. Several methods were used to compare the various outcome measures. For each potential outcome measure, we determined 1) the 95% confidence interval around the observed difference in response rates between treatment and control; 2) the *P* value for between-group differences that would have been achieved (by Fisher's 2-tailed exact test) if that measure had been prespecified as the primary outcome measure; and 3) the group size that would have been required to achieve a *P* value less than 0.05 based on the actual data from the trial.

RESULTS

Figure 1A shows the results at 52 weeks for the CR, total response (CR plus PR), and MCR in each

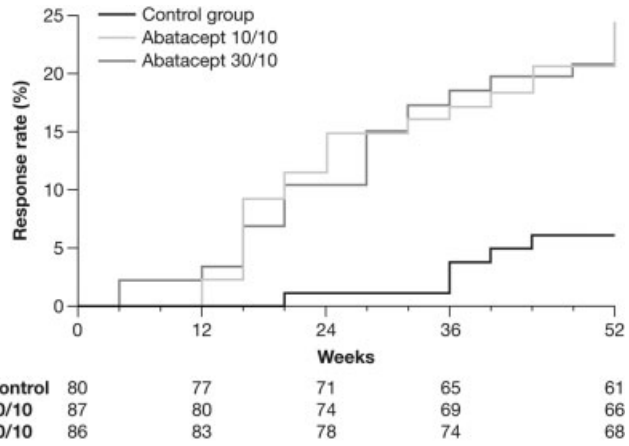

Figure 3. Time to complete response (CR) among patients who met the CR criteria at week 52 in the abatacept treatment groups (as described in Figure 1) compared with the placebo control group. This outcome measure was based on the principle that the concept of time to CR is meaningful only for patients who maintained the CR status and not for those whose CR was evanescent. Therefore, only the data for the patients who met the CR criteria at week 52 are shown. Values below the plot are the number of patients evaluated at each time point.

Table 2. Comparison of alternative outcome measures*

	Response rate, no./total (%)	Difference vs. control (95% CI)	*P* vs. control†
Complete response at 1 year (day 365)			
Control group	5/80 (6)	–	–
Abatacept 10/10	19/87 (22)	16 (5.0, 26)	0.004
Abatacept 30/10	21/86 (24)	18 (7.3, 29)	0.001
Total response at 1 year (day 365)‡			
Control group	22/80 (28)	–	–
Abatacept 10/10	37/87 (43)	15 (−0.5, 29)	0.052
Abatacept 30/10	36/86 (42)	14 (−0.2, 28)	0.073
Major clinical response at 1 year (day 365)			
Control group	11/80 (14)	–	
Abatacept 10/10	29/87 (33)	19 (6.7, 32)	0.004
Abatacept 30/10	31/86 (36)	22 (9.2, 34)	0.001
Treatment failure			
Control group	45/80 (56)	–	
Abatacept 10/10	34/87 (39)	17 (2.0, 31)	0.015
Abatacept 30/10	35/86 (41)	15 (0.4, 30)	0.062
Complete response at week 24 (day 169)			
Control group	4/80 (5)	–	
Abatacept 10/10	15/87 (17)	12 (2.6, 22)	0.017
Abatacept 30/10	11/86 (13)	8 (−1.2, 17)	0.105
Total response at week 24 (day 169)†			
Control group	26/80 (32)	–	
Abatacept 10/10	32/87 (37)	5 (−10, 18)	0.63
Abatacept 30/10	39/86 (45)	13 (−2.0, 27)	0.11
Major clinical response at week 24 (day 169)			
Control group	4/80 (5)	–	
Abatacept 10/10	16/87 (18)	13 (3.6, 23)	0.009
Abatacept 30/10	12/86 (14)	9 (−0.2, 18)	0.066
UPCR improved by ≥75% at 1 year			
Control group	18/80 (22)	–	
Abatacept 10/10	33/87 (38)	16 (1.4, 28)	0.020
Abatacept 30/10	32/86 (37)	15 (0.7, 28)	0.044
eGFR improved by ≥25% at 1 year			
Control group	12/80 (15)	–	
Abatacept 10/10	18/87 (21)	6 (−6.2, 17)	0.42
Abatacept 30/10	16/86 (19)	4 (−8.0, 15)	0.68
Time to complete response	§	§	
Control group			–
Abatacept 10/10			0.002
Abatacept 30/10			0.001

* 95% CI = 95% confidence interval; UPCR = urine protein:creatinine ratio; eGFR = estimated glomerular filtration rate.
† Fisher's exact test was used to compare the treatment groups with the control group for all parameters except time to complete response. The log rank test of equality was used for the analysis of time to complete response.
‡ Defined as complete response plus partial response.
§ See Figure 3 for the data on time to complete response.

abatacept treatment group compared to the placebo control group. The magnitude of the difference between treatment and control was greater for the CR (4-fold) and MCR (2.5-fold) than it was for total response (CR plus PR) (1.5-fold). Figure 1B shows the performance of the same outcome measures at 24 weeks rather than 52 weeks. Similarly, at this earlier time point, the magnitude of the difference between treatment and control was greater for the CR and MCR (3- to 4-fold) than it was for total response (CR plus PR) (<1.5-fold).

Figure 2 shows the results at week 52 for improvement in proteinuria (≥75% reduction in the UPCR), improvement in the eGFR (≥25% increase), and treatment failure. For each of these measures, the treatment groups fared better than the control group, but the magnitude of the difference between treatment and control was <2-fold. Much more striking was the result for time to CR (Figure 3), which was considerably shorter in the treatment groups.

We utilized several approaches to compare the relative performance of the various outcome measures. Table 2 shows the response rate data for each outcome

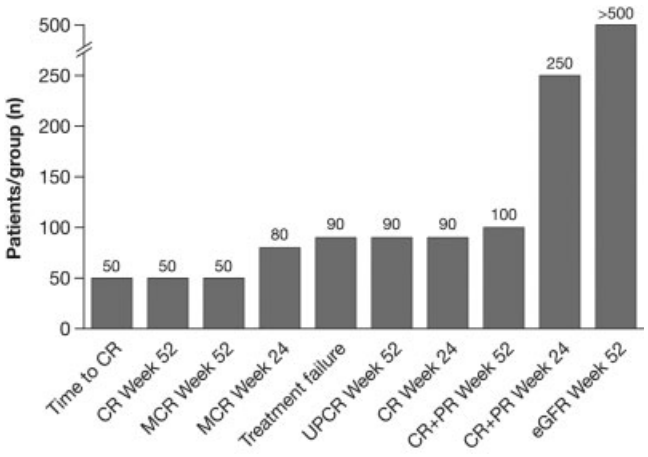

Figure 4. Group sizes representing the approximate number of patients per group that would have been sufficient to demonstrate a statistically significant difference between treatment and control, based on the observed response rates. For the 2 treatment groups, the mean value was used. Values for approximated group sizes are shown over the bars. CR = complete response; MCR = major clinical response; UPCR = urine protein:creatinine ratio; PR = partial response; eGFR = estimated glomerular filtration rate.

measure, along with the 95% confidence intervals around the actual difference in response rates between each treatment group and the control group. Table 2 also shows the *P* value that would correspond to these differences if a particular measure had been prespecified as the primary outcome measure. Figure 4 shows the group sizes that would have been required to demonstrate a statistically significant difference between treatment and control if the measure had, in fact, been the prespecified outcome measure. By applying each of these approaches, we found that the CR at week 52, the MCR at week 52, and the time to CR were the primary outcome measures that best discriminated between treatment and control. The *P* values for comparison of both abatacept treatment groups with the control group were <0.004 for these 3 measures, and the 95% confidence intervals around the response rates for the 3 measures most clearly distinguished treatment from control.

For each of these 3 primary outcome measures, group sizes of 50 patients would have been sufficient to demonstrate a statistically significant difference based on the observed response rates in the treatment and control groups. The other outcome measures were distinctly less effective at discriminating treatment from control. Their order, from most to least discriminating, and the approximate number of patients that would have been required to show statistical significance, was as follows: MCR at week 24 (80 per group), treatment failure at week 52 (90 per group), UPCR reduced by $\geq$75% at week 52 (90 per group), CR at week 24 (90 per group), total response at week 52 (100 per group), total response at week 24 (250 per group), and eGFR increased by $\geq$25% at week 52 (>500 per group) (Figure 4).

DISCUSSION

The challenge of developing effective therapies for lupus nephritis is underscored by the fact that no drug has ever been approved for this indication. This challenge is compounded by the lack of direct evidence to guide us in determining what outcome measure might be most sensitive in detecting differences between treatment groups. The findings presented here suggest 2 tentative conclusions with regard to trial design. First, at least for the data that we evaluated, the implication is that CR rates are preferable to alternative outcome measures that have been used in previous lupus nephritis trials. Second, depending on the size of the trial, 6 months may not be sufficient to convincingly demonstrate efficacy; 1 year is likely to provide a clearer distinction between treatment and control.

The results of our analyses, and the conclusions summarized above, should be viewed with caution. These findings reflect the results of a single trial involving a single agent. Moreover, while additional trials of abatacept in lupus nephritis are under way, the results of this trial have yet to be replicated. Even if subsequent studies of abatacept confirm these results, it is possible that other agents might work more or less quickly, yield a different distribution between complete and partial responses, or even work through mechanisms that require different response criteria. Therefore, these analyses should be viewed as only one step toward building a stronger evidence base for the design of lupus nephritis trials.

Although it has become part of the lupus nephritis lexicon to refer to "complete" responses, it is not possible to distinguish clinically between true quiescence within the kidney and underlying subclinical activity. It is also the case that some abnormalities in renal function may be irreversible, precluding achievement of current definitions of CR; nonetheless, ongoing inflammation in the kidney may have been eradicated. Thus, concepts such as "complete response" or remission may not accurately reflect the true disease state. For this reason, we explored an alternative outcome measure, the MCR, which was designed to include the greatest clinical

responses, without making assumptions about completeness. This outcome measure included not only patients who had met the CR criteria, but also patients who had nephrotic-range levels of proteinuria at baseline and whose UPCR subsequently improved to <1. Although this outcome measure performed as well as the CR definition, it did not significantly improve it.

Although the CR (or MCR) rates appear to be most sensitive in discriminating between groups in the short term (over 1 year), it remains to be determined whether this measure will correlate best with long-term outcome. In the final analysis, the goal is not achieved at 1 year in patients with lupus nephritis. Rather, it is achieved over many years, by preventing progressive renal insufficiency and end-stage renal disease. Past studies that focused on treatment failure as the outcome measure, such as the National Institutes of Health and Euro-Lupus trials of cyclophosphamide, chose a longer trial duration as a means of achieving a more compelling end point. The reality of drug-development timelines push us toward shorter trial designs and to suitable end points that fit these designs. Moreover, patients and physicians depend heavily on early measures of response to guide therapeutic decisions.

Nonetheless, long-term patient outcomes are the ones that matter most. In the absence of data on the long-term outcome among the patients in this trial, we cannot determine whether the superior sensitivity of the CR outcome measure will come at the expense of specificity in predicting long-term outcome. In this regard, it is reassuring that recent studies have shown a strong correlation between complete response and eventual patient and renal survival, as well as a correlation between poor early responses and poor patient and renal outcomes (8). These studies suggest that complete responses not only may be more sensitive than other outcome measures in trials of modest length, but also may be a better measure for predicting long-term outcome.

ACKNOWLEDGMENTS

We thank Stephanie L. Meadows-Shropshire of Bristol-Myers Squibb (Princeton, NJ) for providing help with the statistical analysis. Editorial assistance was provided by Sarah Funderburk, PhD, of Caudex Medical, funded by Bristol-Myers Squibb.

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. Wofsy 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. Wofsy, Hillson, Diamond.
Acquisition of data. Wofsy, Hillson, Diamond.
Analysis and interpretation of data. Wofsy, Hillson, Diamond.

ADDITIONAL DISCLOSURES

Jan L. Hillson, MD is an employee of Bristol-Myers Squibb and was involved in the study design, analysis and interpretation of the data, and the writing of the manuscript. Bristol-Myers Squibb had no role in the decision to submit the manuscript for publication, and publication of this article was not contingent upon approval by Bristol-Myers Squibb.

REFERENCES

1. Wofsy D, Hillson JL, Diamond B. Abatacept for lupus nephritis: alternative definitions of complete response support conflicting conclusions. Arthritis Rheum 2012;64:3660–5.
2. Austin HA III, Klippel JH, Balow JE, le Riche NG, Steinberg AD, Plotz PH, et al. Therapy of lupus nephritis: controlled trial of prednisone and cytotoxic drugs. N Engl J Med 1986;314:614–9.
3. Houssiau FA, Vasconcelos C, D'Cruz D, Sebastiani GD, de Ramon Garrido E, Danieli MG, et al. Immunosuppressive therapy in lupus nephritis: the Euro-Lupus Nephritis Trial, a randomized trial of low-dose versus high-dose intravenous cyclophosphamide. Arthritis Rheum 2002;46:2121–31.
4. Appel GB, Contreras G, Dooley MA, Ginzler EM, Isenberg D, Jayne D, et al. Mycophenolate mofetil versus cyclophosphamide for induction treatment of lupus nephritis. J Am Soc Nephrol 2009;20:1103–12.
5. Rovin BH, Furie R, Latinis K, Looney RJ, Fervenza FC, Sanchez-Guerrero J, et al, for the LUNAR Investigator Group. 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.
6. Tan EM, Cohen AS, Fries JF, Masi AT, McShane DJ, Rothfield NF, et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 1982;25:1271–7.
7. Houssiau FA, Vasconcelos C, D'Cruz D, Sebastiani GD, de Ramon Garrido E, Danieli MG, et al. The 10-year follow-up data of the Euro-Lupus Nephritis Trial comparing low-dose and high-dose intravenous cyclophosphamide. Ann Rheum Dis 2010;69:61–4.
8. Chen YE, Korbet SM, Katz RS, Schwartz MM, Lewis EJ. Value of a complete or partial remission in severe lupus nephritis. Clin J Am Soc Nephrol 2008;3:46–53.