On March 10, 2011, the FDA approved the monoclonal antibody belimumab for the treatment of autoantibody positive systemic lupus which is unresponsive to standard of care treatments. Belimumab is a finely targeted biologic that interferes with the signaling of a single molecule, BLys (also known as BAFF), which functions in a healthy immune response to prolong survival and promote differentiation of B cells. The biologic concept of interfering with B cell survival and differentiation for lupus, a classic disease of B cell hyperactivity, is obviously sound and belimumab met its primary endpoint and several secondary endpoints in two pivotal clinical trials. Furthermore, long-term follow-up of several hundred patients from early phase studies substantiated the impression obtained from the 52 week and 76 week Phase III studies, which had included more than 1,500 patients around the world. Belimumab, even when added to background immune suppressants with known toxicities, is surprisingly safe and well-tolerated. The approval of belimumab seemed to end the 50-year drought of new approved therapies for lupus, and a number of companies have since entered the field, proposing to invest extraordinary sums of money to test more novel, targeted biologic treatments in similar large, international trials using the successful belimumab formula. This may be problematic.

Belimumab was approved with modest but convincingly persistent efficacy margins in the first pivotal trial. It was also approved despite the fact that in the second pivotal trial only a transient statistical difference was achieved between its highest studied dose (10 mg/kg) and placebo on only one single milestone visit, which happened to be the primary endpoint date. With a 9% difference in responder numbers on that single date, the clinical significance of this marginal and temporary result remains uncertain. For this reason, the specific modifiers that the FDA chose to include in the approval wording may be important. Belimumab is only approved for autoantibody positive patients, and it is only approved for patients unresponsive to standard of care. What does this mean for further treatment development for lupus?

The answer requires a good understanding of history, without which those who plan new drug development are doomed to repeat it with no guarantee that the final outcome at the FDA approval process will be identical to the first drug brought forward with modest efficacy margins, especially since it will not be easy to duplicate the safety record of belimumab.

Limitation of Belimumab Approval to Autoantibody Positive Patients

Some history underlying the integration of “seropositivity” into the SLE drug approval process dates back to two serious problems in the field of lupus treatment development. The first is a generic problem, that of quality assurance in large international clinical trials and the need to ensure appropriate recruitment of patients. There has been a longstanding annoyance among rheumatologists that ANA testing does not diagnose lupus and many patients are inappropriately referred on the basis of this test with no clinical features of concern. In response to this, many large clinical laboratories have increased the stringency of ANA testing in order to eliminate “false positive” results, meaning those people with positive ANA who do not have lupus. Thus, over time, the ANA testing has become increasingly specific to lupus increasing the likelihood that it has lost sensitivity in detecting...
patients with true SLE. Rheumatologists were trained and have long experience to justify the belief that ANA is a sensitive and not very specific test, and therefore may have some trouble interpreting negative ANA in patients who otherwise meet four major classification criteria for SLE. This issue is amplified by the known lack of consistency in autoantibody testing from place to place.9 Thus, an increasingly common event in the past decade has been entry of patients into SLE studies who might have a positive ANA in the local lab but are designated as negative by the central testing laboratory. Finally, more aggressive treatments over time may have led to an increasing percentage of patients who really do lose their ANA positivity. For the clinical classification of SLE, it has not been considered necessary for a patient to always have a positive ANA in repeated testing, and this is not a test that is frequently repeated over time in clinical practice. For all of these reasons, it would have to have been a surprise to many legitimate trialists how frequently their patients with confirmed SLE tested ANA negative when screening for clinical trials. It certainly surprised this investigator, and the apparent stringency of central laboratory testing seems to have increased even more in recent years as compared to local laboratories using classic immunofluorescence methods.

The second problem, related to the first, was that belimumab, like most theoretically well-substantiated biologic treatments for SLE, failed to meet any of its primary or secondary endpoints in its initial Phase II trial.10 This led the company to investigate (appropriately) whether an exploratory analysis might suggest a subgroup or endpoints suitable to justify continued drug development. In that exploratory analysis, which led to a successful redesign for the Phase III trials, it was found that if patients who had been entered into the Phase II study with negative ANA and negative anti-dsDNA were eliminated from analysis, a difference could be detected between patients receiving belimumab versus placebo using an exploratory outcome (SRI), which also served as a basis for the successful Phase III trial design.

Therefore, the entry criteria for the large Phase III belimumab studies restricted entry to patients with positive ANA or anti-dsDNA. The implication that significant numbers of patients actually exist who test positive for anti-dsDNA and negative for ANA substantiates the argument that central laboratory testing is using an increasingly stringent cutoff for positive ANA, since this would be a rare or even non-existent finding using any classic laboratory testing, despite differences in assay results from place to place. The stringent cutoff for ANA may have an important meaning in defining a subset of patients who might have higher titer ANA in general as being better candidates for B cell targeted therapy than those considered negative at the central laboratory, who nevertheless may have true SLE and a positive ANA at lower titers using more classic fluorescence cutoffs in testing.

It is possible that elimination of patients who tested negative for ANA at baseline resulted in a good outcome. Some patients could have been misdiagnosed with SLE based on the false positive ANA and entered in Phase II. If these patients were subsequently eliminated from the belimumab Phase III trials, this might explain its improved results. However, there is a downside to replacing clinical judgment with an increasingly insensitive laboratory test. Patients with high titer ANAs may include a higher percentage too ill to participate in trials, limiting potential enrollment in the increasingly competitive marketplace of very large Phase III trials. Also, it may be that not every targeted treatment, especially those that do not target B cells, will work best in that subgroup with high ANA. However, a positive ANA at a central laboratory at baseline has now, on the basis of this one exploratory analysis of the Phase II belimumab study, become a non-negotiable entry criteria in most ongoing clinical trials. This is regardless of the mechanism of action of some of these agents.

Meanwhile, the FDA has approved belimumab based on these entry criteria but has designated no time point within which the ANA must have been positive nor what laboratory test must be used to designate ANA. Thus, many patients who tested positive for ANA using older methods are, at least technically, considered candidates for this treatment by the FDA, when they were not allowed to participate in Phase III trials. Even more insidiously, the ANA may simply be serving as an indirect and imprecise marker for circulating BLyS levels. By eliminating ANA negative patients an efficacy signal could be determined in one exploratory analysis of a Phase II trial. Consistent with this, more ANA “positive” (by central laboratory) patients had elevated BLyS than those who were ANA “negative.” However, this does not mean that all ANA “positive patients” had elevated BLyS activity suitable for anti-BLyS treatment, nor does it mean that all ANA “negative” patients were not candidates for the treatment.

**Standard of Care**

It should not be surprising that belimumab, like most treatments studied in lupus, failed to meet any of its primary and secondary efficacy endpoints in Phase II.10 The hurdles in treatment assessment for SLE have been well recognized: this is a disease of particular heterogeneity, and even with the advanced molecular science that has led to targeted biologics with outstanding proof of concept established in animal models and in vitro human studies, the field has not advanced to the point that biologic assays can be validated in advance to optimize patient selection, dose, administration protocols, or surrogate outcomes. Without surrogate biologic outcomes, the community relies on imperfect and inconsistent measurements of disease activity for patients who are quite unpredictable in terms of how often or severely the disease waxes and wanes.11,12 Coupled with an imperative glass ceiling on the percentage of all lupus patients likely to have the requisite pathway activity to even potentially respond to the treatment, the potential
efficacy of any novel, targeted therapy must be limited, and it is no wonder that belimumab has not exceeded 50% efficacy in pivotal trials.5,6

Compounding all of these issues has been enough trepidation about the potential for lupus to become a serious or life threatening illness that it has caused the avoidance of trial designs that use pure placebo groups or even one consistent, common background treatment for all patients. Most Phase II and III trials performed to date for SLE have allowed disparate background treatments in both the treatment and placebo groups, and additionally, given the long length of most of these studies, treatment adjustments have been either mandated or at least allowed during the studies, with minimal knowledge of how each of these various agents might be interfering with (or redundant to) the study agents. There is much published exploratory analysis of these studies strongly suggesting that this trial design confounds analysis of treatment effects.13,18

This potpourri of unapproved, understudied, and biologically not well-characterized background treatments in trials is apparently what the FDA is now referring to as “standard of care,” to which it is being recommended that we add belimumab. By insisting that a patient must fail the standard of care prior to justifying a biologic, both in trial design and in eventual approval language (which might seem laudable from both a clinical and an economic point of view), future trials may be doomed to enroll patients on this same variegated array of potent treatments, which (unless the will to stop them as part of a protocol develops) will continue to seriously confound trial outcomes. Why is our current addition of test treatments to a cacophony of poorly characterized treatments (e.g., standard of care) preferable to a trial with a well-defined, biologically characterized, (e.g., rational) comparator? The problem seems to stem from a regulatory mandate to avoid studies in which treatments are compared to equally efficacious but unapproved drugs; something that might be highly acceptable to the community and more efficient for the companies that lose substantial investment when recruitment takes a long time, since in a comparator trial design there would be treatment initiation for all patients without enforced continuation of immune suppressants that are not working when patients enter the trial.

However, if such a trial design resulted in similar efficacy between an investigational agent and a known, effective, “standard of care” immune suppressant, even if the biologic was found to be safer, this would not warrant FDA approval. The FDA must consider all treatments that are not already approved for an indication to be placebo, and superior efficacy to placebo must be demonstrated for approval. Setting aside the quixotic notion that these same treatments are elevated to standard of care status after approval, it seems remarkable that the potential for improved safety to standard of care would not count if efficacy were equal to what is currently being widely used and, indeed, to some extent mandated by the FDA to be used with belimumab. This puts lupus development into a Catch-22 situation in which patients must receive adequate treatment in these year-long pivotal trials, with a biologic ceiling on the potential efficacy of any single-targeted agent, and the need to ensure almost that much efficacy for the safety of the placebo group, making it quite difficult to demonstrate superiority of any treatment. Is it any wonder, then, that the results of the belimumab Phase III pivotal trials were somewhat anemic?

Table 1 summarizes the results of the belimumab Phase III studies, which were performed using identical study design in different geographic regions.5,6 What is immediately obvious is that, despite differences in the numbers of patients who met endpoints overall from trial to trial, the ability to assess the meaning of belimumab responses in both studies is confounded by the relatively high rates of response in the placebo groups, which were of course patients receiving standard of care and in some cases aggressive standard of care. In both trials, patients in all groups could have immune suppressant doses increased for the first 16 weeks and increased steroids for the first six months.5,6 Previous studies have shown that when steroids are added to standard of care but later tapered, many patients who were otherwise treated with placebo will continue to do well for another 6 months until the end of a year-long trial.14,15

Table 2 summarizes the treatments that patients in the Phase III trial were taking at the beginning of the study. In the BLISS-52 study, 96% of the patients were taking some steroids at entry (69% > 7.5 mg), and 42% were taking immune suppressive treatments. In the BLISS-76 study, 76% were on steroids (46% > 7.5 mg), and 56% were on immune

Table 1  BLISS 52 and BLISS 76: 52-Week Results5,6

<table>
<thead>
<tr>
<th>Primary Endpoint</th>
<th>Treatment</th>
<th>Placebo</th>
<th>1 mg/kg</th>
<th>10 mg/kg</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLISS 52 (865 pts)</td>
<td>Placebo</td>
<td>43.6</td>
<td>51.7 (p = 0.011)</td>
<td>57.6 (p = 0.0006)</td>
</tr>
<tr>
<td>BLISS 76 (819 pts)</td>
<td>Placebo</td>
<td>33.8</td>
<td>40.6 (p = 0.1)</td>
<td>43.2 (p = 0.021)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Secondary Endpoints</th>
<th>Treatment</th>
<th>Placebo</th>
<th>1 mg/kg</th>
<th>10 mg/kg</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLEDAI improvement</td>
<td>Placebo</td>
<td>46.0</td>
<td>53.1 (p = 0.019)</td>
<td>58.3 (p = 0.0024)</td>
</tr>
<tr>
<td>Prednisone reduction*</td>
<td>Placebo</td>
<td>12.0</td>
<td>20.6 (p = 0.025)</td>
<td>18.6 (p = 0.053)</td>
</tr>
</tbody>
</table>

*25% below baseline dose to ≤ 7.5 mg per day during weeks 40 to 52. This analysis could be done on 69.4% of patients in BLISS 52 and 45% of patients in BLISS 76.
Data that are not yet evaluated but that might be critical to the design of future clinical trials, is how many patients in each group had increased treatment during the first half of the study. Steroid tapering at the end of the studies has, however, provided some interesting data (Table 1). During weeks 46 to 52, among those patients taking > 7.5 mg/day of prednisone or equivalent doses of other steroids at entry, more patients treated with belimumab, as compared to placebo, were able to decrease these doses by at least 25% of the entry dose and to less than 7.5 mg/day prednisone or equivalent. This reached statistical significance in the BLISS-52 study, but a similar trend was present in the BLISS-76 study, which had fewer patients and a lower percentage of the patients were eligible for this analysis, thus lowering the statistical power for the evaluation in that trial.

How would this trend affect the endpoints? Since more patients achieved lower doses of steroids in the belimumab treatment groups in both trials, there could be either no effect on the endpoints or it could narrow the gap between treatment and placebo; however, there is no rationale to assume this ambitious “add on” part of the trial design helped to distinguish the treatment effects from placebo in the primary endpoint. Is an extra steroid-sparing analysis worth adding to a trial design? Do rheumatologists need this potentially confounding component in an efficacy trial design to assume that any effective treatment must be steroid sparing? This seems unlikely. What is still not known is the range of doses in those patients taking more than 7.5 mg prednisone or equivalent at entry. This might enable a comparison by how much actual dose lowering was taking place and allow a better estimation of the impact steroid tapering might have had on the outcomes of each trial.

What Were the Differences Between the Two Phase III Trials of Belimumab?

Table 1 compares the primary endpoints at week 52 in the two main, pivotal trials, but the major difference between these two studies was an early and consistent demonstration of belimumab efficacy in the BLISS-52 study and only a transient effect, which was statistically significant in the higher dosing group at week 52 alone in BLISS-76.5,6 What are the possible reasons for the striking differences in outcomes between these studies?

<table>
<thead>
<tr>
<th>Medications</th>
<th>BLISS 52 (N = 865)</th>
<th>BLISS 76 (N = 819)</th>
<th>Increase in Meds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steroids &gt; 7.5 mg pred/equiv</td>
<td>69%</td>
<td>46%</td>
<td>Allowed</td>
</tr>
<tr>
<td>IS (total)</td>
<td>42%</td>
<td>56%</td>
<td>Allowed</td>
</tr>
<tr>
<td>AZA</td>
<td>26%</td>
<td>20%</td>
<td>Allowed</td>
</tr>
<tr>
<td>MTX</td>
<td>9%</td>
<td>18%</td>
<td>Allowed</td>
</tr>
<tr>
<td>MMF</td>
<td>6%</td>
<td>15%</td>
<td>Allowed</td>
</tr>
<tr>
<td>Antimalarial</td>
<td>67%</td>
<td>63%</td>
<td></td>
</tr>
<tr>
<td>NSAIDs</td>
<td>41%</td>
<td>20%</td>
<td></td>
</tr>
</tbody>
</table>

FDA Briefing Package, Badrul A. Chowdhury, October 19, 2010.

Given the modest effect sizes in both trials, it is possible that no reason is needed to explain the inconsistent and marginal proof of efficacy found in BLISS-76 compared to the modest but completely consistent dose-responsive efficacy demonstrated in BLISS-52. However, the following circumstantial difference between the two studies remains to be considered: BLISS-76 took place in Western Europe, the United States, Canada, and Mexico. BLISS-52 took place in South America, Eastern Europe, Asia, India, and Australia.5,6

Geography could be reflective of different environments but might also affect the racial (and indirectly the genetic) background of the patients, the probability that the BLyS pathway was relevant in a substantial number of subjects, and the choice or dosing of background immune suppressants, as well as the likelihood of aggressive adjustments as allowed within the ongoing protocol. Other than the differences in background treatments at entry into the trial reviewed above, very little is known about these potential differences. Attempts to consider some of these issues with exploratory analyses by country or race were made for the FDA discussions,19 but little can be understood from analyses of small subsets of large studies with modest effect sizes, a conclusion that was correctly drawn by FDA statisticians.19 For this reason, ongoing studies will be needed to address some of these issues, especially some uncertainties regarding efficacy in patients of African descent and patients with renal disease.

Conclusions

Many issues may be confounding to clinical trials, but it stands to reason that better biologic understanding of disease subsets and trial designs that allow improved discrimination between new treatments and placebo by minimizing the potentially confounding effects of background medications would provide the optimal path forward in assurance of interpretable clinical trials and improved care for patients with lupus.

Disclosure Statement

The author has no financial or proprietary interest in the subject matter or materials discussed, including, but not limited to, employment, consultancies, stock ownership,
honoraria, and paid expert testimony.

References