ACR Open Rheumatology
Vol. 4, No. 10, October 2022, pp 923–930
DOI 10.1002/acr2.11451

American College of Rheumatology
Empowering Rheumatology Professionals

External Validation of the Lupus Multivariable Outcome Score for Systemic Lupus Erythematosus Trials

Michal Abrahamowicz,[1] Maria Izabela Abrahamowicz, and Peter E. Lipsky[2]

Objective. Development of new systemic lupus erythematosus (SLE) treatments requires an effective responder index. Toward this end, we have recently developed a new Lupus Multivariable Outcome Score (LuMOS) to optimize discrimination between actively treated patients and those on placebo. We now report on external validation of LuMOS in two independent clinical trials.

Methods. Validation was performed with the Illuminate data sets that evaluated tabalumab (TB) in SLE. To accommodate laboratory results assessed on different platforms, we developed a standardized LuMOS 2.0 model that uses *z* score transformations of biomarker values. For validation, we calculated LuMOS 2.0 scores at week 52 for all participants. Effect size (ES), with 95% confidence intervals (CIs), compared the ability of LuMOS and the SLE Responder Index-5 (SRI-5) to discriminate between outcomes in patients randomized to TB dosage and outcomes in those randomized to a placebo.

Results. Mean LuMOS 2.0 scores were significantly higher ($P < 0.0001$) for the TB groups than the placebo group, including the Illuminate-1 trial, in which the SRI-5 did not identify significant treatment effects. For both TB groups in both trials, LuMOS 2.0–based ES indicated moderately strong treatment effects (>0.4) in contrast to weak SRI-5 effects (<0.25). For monthly TB, LuMOS 2.0–based ES were 0.44 (95% CI: 0.30-0.59) and 0.54 (95% CI: 0.39-0.68) for the Illuminate-1 and Illuminate-2 trials versus corresponding SRI-5-based ES of 0.13 (95% CI: −0.02 to +0.27) and 0.15 (95% CI: 0.01-0.30).

Conclusion. LuMOS 2.0 detected significantly greater treatment effects compared with the SRI-5 in the Illuminate trials. Additional validation of LuMOS 2.0 in trials of non–B cell–directed therapies will be necessary to document its universality as an outcome measure.

INTRODUCTION

A validated responder index that is responsive to clinically relevant changes would increase the ability of randomized controlled trials (RCTs) to identify effective new treatments in systemic lupus erythematosus (SLE) (1–5). One important reason for the problematic nature of outcome measures in SLE is the multisystem nature of the disease and the desire to capture all the clinical manifestations in a single metric. Because there are considerable differences in the response of individual patients enrolled in RCTs to a specific treatment, capturing the diversity of outcomes has posed a challenge. To address this complexity, most SLE responder indices attempt to aggregate information on potential treatment-induced improvements in disparate relevant variables, including both clinical disease activity and laboratory measurements (6–11). Different outcome measures have been proposed using different methods and criteria, but head-to-head comparisons of their responsiveness to change are limited (5). Furthermore, only a few studies reported systematic external validation of the more frequently employed SLE responder indices or composite outcomes in independent trials, and the results were sometimes negative. For example, in a recent validation study, a new SLE responder index performed well in the training sample used to develop the index but did not discriminate accurately between responders and nonresponders in an independent validation sample (3).

To address the aforementioned challenges, we have recently developed a new Lupus Multivariable Outcome Score

[1]Michal Abrahamowicz, PhD: McGill University and Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada; Maria Izabela Abrahamowicz, MSc; [2]Peter E. Lipsky, MD: AMPEL BioSolutions and Re-Imagine Lupus Investigation, Treatment and Education (RILITE) Research Institute, Charlottesville, Virginia.

Author disclosures are available at https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Facr2.11451&file=acr211451-sup-0001-Disclosureform.pdf.

Address correspondence to Peter E Lipsky, MD, Re-Imagine Lupus Investigation, Treatment and Education (RILITE) Research Institute, Suite 300, 250 West Main Street, Charlottesville, VA 22902. Email: peterlipsky@comcast.net.

Submitted for publication November 3, 2021; accepted in revised form March 25, 2022.

(LuMOS) (12). The LuMOS responder index was derived through multivariable regression analyses of the successful Study of Belimumab in Subjects with SLE 76-week (BLISS-76) trial (13), which, together with a similarly designed BLISS-52 trial (14), led to approval of belimumab (Bel) for SLE. The LuMOS score was validated in three analyses of either BLISS-76 (13) or BLISS-52 (14) data, which compared patients randomized to active Bel treatment versus a placebo. In all three validation analyses, mean LuMOS scores for patients treated with Bel were significantly higher ($P < 0.0001$) than those for the placebo group, and LuMOS yielded much higher values of Cohen's *d* effect size (ES) (15) than the corresponding ES obtained with SLE Responder Index-4 (SRI-4) (12).

To evaluate the performance of LuMOS in SLE trials more completely, we now report on further validation of LuMOS in independent RCTs after modifications of the scoring system to make its application more universal. Specifically, we first developed a modified standardized LuMOS score, which can be used regardless of the potential between-trial differences in the assays employed to measure C3, C4, or anti–double-stranded DNA (dsDNA) titer, and then validated the resulting standardized LuMOS (LuMOS 2.0) using data from two independent SLE trials, Illuminate-1 and Illuminate-2, that evaluated tabalumab (TB) (16,17).

PATIENTS AND METHODS

Data sources. Data for the BLISS-52 (13) and BLISS-76 (14) trials were obtained from GlaxoSmithKline through their open data access program. Both trials randomized participants to three arms: 1) high-dose Bel, 10 mg/kg; 2) low-dose Bel, 1 mg/kg; and 3) placebo. Data for the Illuminate-1 (10) and Illuminate-2 (11) trials were obtained from Eli Lilly, Co. Both Illuminate trials included three randomization groups: 1) higher-frequency TB (every 2 weeks), 2) lower-frequency TB (every 4 weeks), and 3) placebo (16,17). In all four trials, the background was standard of care medications. Study design and inclusion and exclusion criteria are reported in the original publications (13,14,16,17) and are summarized in section 1 of the Supplementary Material. Similar to the original primary analyses, we relied on intention-to-treat analyses that included all originally randomized participants, with last observation carried forward for all missing values (13,14,16,17).

Development of the standardized LuMOS 2.0 score to address variation in the assays. The original LuMOS score (12) was developed using the data from the BLISS-76 trial (13). Accordingly, for laboratory markers (C3 and C4 components and anti-dsDNA titer) the estimated LuMOS coefficients reflected the corresponding changes measured using the specific enzyme-linked immunosorbent assays employed in that particular study (13). However, different studies often employ different assays, resulting in nonnegligible differences between the measurements of the same variable. To develop a revised formula, robust with respect to the use of different assays, we have considered two alternative approaches to standardization of the values of the three biomarkers: 1) the relative change LuMOS 1.2 model and 2) the standardized change LuMOS 2.0 model.

The LuMOS 1.2 model relied on the relative change, for each study subject, calculated by dividing the difference in the biomarker values observed at week 52 and week 0 by the initial (week 0) value. The resulting relative changes are independent of the assays and/or measurement units used in a given trial.

LuMOS 2.0 employs standardized change in the biomarker values. To this end, the original biomarker value, *x(i)*, observed for subject *i* at a given assessment was replaced by a standardized value using *z* score transformation:

$$z(i) = [x(i) - m]/sd,$$

where *m* is the mean of all relevant baseline (week 0) values across all study subjects, and *sd* is their standard deviation. Biomarkers were thus expressed using generic *z* score units, with one unit corresponding to the respective standard deviation, and were independent of the measurement units and assays used. Next, for each subject, we calculated the standardized change in a specific biomarker as a difference between the corresponding *z* score at week 52 and the one at week 0.

The resulting *z* score differences for each biomarker (C3 and C4 components and anti-dsDNA titer) were then used, together with the original untransformed changes in the other predictors included in the original LuMOS model (12), to reestimate the corresponding new LuMOS 2.0 model. Specifically, the LuMOS 2.0 scoring formula was developed using multivariable logistic regression with the binary outcome defined as the contrast (at 52 weeks) between the Bel-10 mg/kg group and the placebo group in the BLISS-76 trial (13) (ie, the same contrast used to develop the original LuMOS model) (12).

Next, we applied the resulting LuMOS 1.2 and LuMOS 2.0 formulas to the independent validation data set from the BLISS-52 trial (14). Here, for the *z* score standardization required by LuMOS 2.0, we used the respective mean and standard deviation values from the BLISS-52 trial (while using regression coefficients estimated from the BLISS-76 analyses). Then, to assess the ability of the new LuMOS 1.2 and LuMOS 2.0 models to discriminate between the outcomes for actively treated subjects and the outcomes for those randomized to the placebo, for each treatment group, we estimated the ES (15), relative to the placebo group. The ES was calculated as the ratio of the difference of the mean LuMOS score in the active treatment group minus the mean LuMOS score in the placebo group (numerator) divided by the pooled within-group standard deviation of the scores (denominator) (15), and was reported together with 95% confidence intervals (CIs) (18,19). Finally, we compared the resulting ES values with the corresponding ES for the original unstandardized LuMOS model (12) applied to the same BLISS-52 data.

External validation of the standardized LuMOS 2.0 model in two Illuminate trials of TB. To provide evidence of external validity of the standardized LuMOS 2.0 model as a SLE responder index, we assessed its ability to discriminate between the outcomes in the placebo group and the outcomes in each of the two active treatment groups in two independent randomized trials (Illuminate-1 and Illuminate-2) that evaluated TB treatment in SLE (16,17).

To this end, we first used the standardized scoring formula in Table 2 to calculate LuMOS 2.0 scores for each study participant, based on changes in the relevant variables observed between randomization [week 0] and week 52, used to assess outcomes in the primary analyses of both trials (16,17). Next, we compared the mean LuMOS 2.0 scores and the corresponding standard deviations between the three randomization groups. To quantify the ability of the LuMOS 2.0 scores to discriminate outcomes in patients with SLE treated with TB from outcomes in those who received a placebo, we used the ES (9) with 95% CIs (16,17). If the 95% CI for the ES for a given TB group excluded 0, this indicated that LuMOS 2.0 detected a statistically significant (at two-tailed $\alpha = 0.05$) treatment effect (12). ES values of 0.2, 0.5, and 0.8 are typically interpreted as a small, moderate, and large effect, respectively (15).

Finally, for each between-groups comparison, we also calculated the corresponding ES based on the binary SLE Responder Index-5 (SRI-5) (11) used as the primary outcome measure in both Illuminate trials (16,17). To calculate the ES values, we considered that, for a binary random variable, 1) the group mean equals the proportion of responders (p) and 2) the standard deviation equals $\sqrt{[p(1-p)]}$. We then compared the corresponding ES values, with 95% CIs, for LuMOS 2.0 versus SRI-5 to assess their relative performance in discriminating between outcomes observed during 52 weeks of the trial in patients treated with TB and outcomes in those randomized to a placebo.

Relationships of LuMOS 2.0 and SRI-5 responses with TB treatment. Additional analyses were conducted in each of the two Illuminate data sets to further explore whether and to what extent LuMOS 2.0 scores and/or SRI-5 responses are able to discriminate between patients treated with TB and those who received a placebo. To this end, separately for data from each Illuminate trial, we estimated a multivariable logistic model, with the dichotomous outcome defined as the patient's inclusion in one of the two TB groups versus the placebo group. Independent variables included age and sex, as well as two responder indices: binary SRI-5 response and continuous standardized LuMOS 2.0 score, both evaluated at week 52, as in the original Illuminate publications (16,17). LuMOS 2.0 scores were converted to *z* scores so that the resulting adjusted odds ratio (OR) represented the effect of increasing the score by 1 standard deviation of the distribution of LuMOS scores across all participants in the trial. The results of primary interest were the mutually adjusted ORs for SRI-5 responses and LuMOS 2.0 scores, with 95% CIs, and their statistical significance was tested using model-based Wald χ^2 tests with 1 degree of freedom at two-tailed $\alpha = 0.05$. The results of these tests indicated whether SRI-5 responses and LuMOS 2.0 scores are able to significantly improve discrimination between outcomes of patients treated with TB and outcomes of those randomized to the placebo over the discrimination achieved with the other responder index. Because the OR for a binary SRI-5 cannot be directly compared with the OR for a 1–standard deviation increase in the continuous LuMOS 2.0 score, the relative strength of their adjusted effects was compared based on the corresponding values of the Wald statistics.

In sensitivity analyses, we explored whether the adjusted association between continuous LuMOS 2.0 scores and the logit of the binary outcome (inclusion in one of the TB groups) may be nonlinear. To address this issue, we reestimated a flexible extension of the aforementioned multivariable logistic model using cubic regression splines with 5 degrees of freedom to model possibly nonlinear effect of increasing LuMOS 2.0 scores (20–22). The fit of the flexible nonlinear spline-based model was compared with that of the conventional logistic model using the Akaike Information Criterion (23), with a reduction of at least 4 Akaike Information Criterion points considered as evidence of a clinically important nonlinear effect (24).

RESULTS

Validation of the standardized LuMOS 2.0 and comparison with the original LuMOS. Table 1 compares the performance of the original LuMOS (6), the relative change LuMOS 1.2, and the standardized LuMOS 2.0 models. The performance of each model is quantified by the ES, with higher ES indicating better discrimination (ie, a model more responsive to change) (12,15). A model indicates a statistically significant treatment effect, at two-tailed $\alpha = 0.05$, if the corresponding 95% CIs for the ES (last column) exclude 0 (15,19).

For all the comparisons reported in Table 1, including validation analyses in an independent BLISS-52 data set (14), all three of the LuMOS models provide clear evidence of statistically significant differences in the outcomes observed in each of the Bel groups versus the placebo group. Indeed, all 95% CIs for the ES exclude not only the null effect (ES = 0) but also any very weak effect (ES $\leq$ 0.15) (15). However, there were meaningful differences in the responsiveness to change of the alternative LuMOS models. In particular, the relative change LuMOS 1.2 model performed systematically worse than either the original LuMOS model or the standardized LuMOS 2.0 model for either of the active treatment groups (Table 1). Indeed, the relative change LuMOS 1.2 model yielded ES values lower by 0.17 to 0.19 (ie, by more than 30%, in relative terms) than the corresponding ES

Table 1. Comparing the original versus two new LuMOS models: distributions of scores and ES

Model or trial	Group (n)	Biomarkers values used	Mean score	SD	ES (relative to PB)	95% CI for ES
Original LuMOS						
BLISS-76	PB (n = 275)	Change in original values	−0.22	0.69	–	–
BLISS-76	Bel-1mg (n = 271)	Change in original values	+0.07	0.69	0.43	0.26-0.60
BLISS-76	Bel-10mg (n = 273)	Change in original values	+0.23	0.70	0.66	0.49-0.83
LuMOS 1.2						
BLISS-76	PB (n = 275)	Relative change (%)	−0.17	0.61	–	–
BLISS-76	Bel-1mg (n = 271)	Relative change (%)	+0.04	0.59	0.35	0.18-0.52
BLISS-76	Bel-10mg (n = 273)	Relative change (%)	+0.04	0.59	0.53	0.36-0.70
LuMOS 2.0						
BLISS-76	PB (n = 275)	Change in *z* scores	−0.22	0.69	–	–
BLISS-76	Bel-1mg (n = 271)	Change in *z* scores	+0.07	0.69	0.42	0.25-0.59
BLISS-76	Bel-10mg (n = 273)	Change in *z* scores	+0.23	0.70	0.64	0.47-0.81
External validation in BLISS-52						
Original LuMOS model						
BLISS-52	PB (n = 287)	Change in original values	−0.28	0.75	–	–
BLISS-52	Bel-1mg (n = 288)	Change in original values	+0.16	0.78	0.57	0.40-0.74
BLISS-52	Bel-10mg (n = 290)	Change in original values	+0.29	0.77	0.75	0.58-0.92
LuMOS 1.2						
BLISS-52	PB (n = 287)	Relative change (%)	−0.18	0.60	–	–
BLISS-52	Bel-1mg (n = 288)	Relative change (%)	+0.06	0.59	0.40	0.23-0.57
BLISS-52	Bel-10mg (n = 290)	Relative change (%)	+0.19	0.71	0.56	0.39-0.72
LuMOS 2.0						
BLISS-52	PB (n = 287)	Change in *z* scores	−0.28	0.75	–	–
BLISS-52	Bel-1mg (n = 288)	Change in *z* scores	+0.16	0.78	0.57	0.40-0.74
BLISS-52	Bel-10mg (n = 290)	Change in *z* scores	+0.29	0.77	0.74	0.57-0.91

Abbreviations: Bel-1mg, belimumab 1 mg/kg; Bel-10mg, belimumab 10 mg/kg; BLISS-52, Study of Belimumab in Subjects with SLE 52-week; BLISS-76, Study of Belimumab in Subjects with SLE 76-week; CI, confidence interval; ES, effect size; LuMOS, Lupus Multivariable Outcome Score; PB, placebo; SD, within-group standard deviation.

values obtained with the two other LuMOS models. In contrast, the mathematical properties of the standardized change LuMOS 2.0 model ensure that it performs as well as the original LuMOS model in all analyses reported in Table 1.

Based on the clear pattern of results presented in Table 1, the new LuMOS 2.0 model, which relies on *z* score standardization of the observed values of each continuous biomarker, is the most effective generalizable version of the LuMOS model. This model can be used regardless of the potential between-study differences in the assays used to measure the biomarkers, units used, and/or distributions of the observed biomarker values.

Table 2 compares the resulting new generalizable scoring formula of the proposed standardized LuMOS 2.0 model with the original LuMOS formula reported previously (12). As expected, the differences in the corresponding coefficients concern mainly the three biomarkers, whereas for all other variables the coefficients are very similar. To further facilitate the implementation of the standardized LuMOS 2.0 model in future SLE studies, section 3 of the Supplementary Material provides a detailed step-by-step description of all the required data manipulations and calculations. These steps are illustrated in the Supplementary Table A.1 using data for a single trial participant.

Validation of LuMOS 2.0 in the Illuminate-1 and Illuminate-2 trials of TB. Table 3 compares the standardized LuMOS 2.0 scores in the three randomization groups at week 52 separately for each of the two Illuminate trials. In both trials, the mean LuMOS 2.0 scores for patients treated with either TB

Table 2. Comparing the LuMOS scoring formulas: original LuMOS model (12) versus the LuMOS 2.0 model

Variable	Coefficient (original LuMOS)	SE (original LuMOS)	Coefficient *z* score (LuMOS 2.0)	SE *z* score (LuMOS 2.0)
Intercept	−0.43	0.142	−0.433	0.14
SLEDAI score decrease ≥4	0.24	0.184	0.238	0.18
Prednisone dose change (1 mg/day)	−0.005	0.004	−0.005	0.004
Anti-dsDNA change	−0.0068	0.0026	−0.506	0.199
C3 change	−0.0002	0.0007	−0.067	0.225
C4 change	0.105	0.0276	1.016	0.266
BILAG renal worsening	−0.39	0.276	−0.392	0.276
BILAG mucocutaneous improvement	0.53	0.302	0.526	0.302

Abbreviations: BILAG, British Isles Lupus Assessment Group and Mucocutaneous; LuMOS, Lupus Multivariable Outcome Score; SE, standard error of the corresponding regression coefficient; SLEDAI, Systemic Lupus Erythematosus Disease Activity Index.

dose are positive, in contrast to negative mean scores for both placebo groups. The responsiveness to change of LuMOS 2.0 is further underscored by the fact that the corresponding 95% CIs for the mean scores for the active treatment groups do not overlap at all with the corresponding 95% CIs for patients randomized to the placebo group (Table 3). In contrast, the between-groups differences in the proportions of SRI-5 responders are considerably less pronounced, with an overlap between 95% CIs for either TB group and the corresponding 95% CI for the placebo group (Table 3). Notably, the overlap is only minimal for the high-frequency (every 2 weeks) TB group in the Illuminate-2 trial, consistent with a statistically significant difference found for this group in the original analyses based on SRI-5 (17).

Table 4 reports the ES values for contrasts between each of the TB groups and the placebo group. For all four contrasts, LuMOS 2.0 yields ES values close to or higher than 0.5 (Table 4), considered a moderate effect (15). Furthermore, all four corresponding 95% CIs exclude not only 0, indicating statistically significant TB effects, but also any ES values close to 0.2, considered a weak effect. In contrast, the SRI-5 (11) indicates much lower ES, three of which are below the 0.2 cutoff for a weak effect (15). In addition, for both TB groups in the Illuminate-1 trial, the 95% CIs for the ES include 0 (Table 4) and, thus, fail to demonstrate a statistically significant treatment effect. Finally, for all four contrasts, even the lower bounds of the 95% CIs for LuMOS 2.0–based ES are higher than the upper bounds of the corresponding 95% CIs for SRI-5–based ES (Table 4). This underscores the finding that, relative to SRI-5, LuMOS 2.0 yields systematically statistically significantly better discrimination between outcomes of patients treated with TB and outcomes of those randomized to a placebo.

Assessing the capacity of LuMOS 2.0 and SRI-5 to discriminate between outcomes in the active treatment groups and outcomes in the placebo group. In both Illuminate data sets, multivariable logistic regression yielded very significant associations between a higher LuMOS 2.0 score at week 52 and the odds of a patient being treated with TB rather than receiving a placebo. Spline-based sensitivity analyses did not reveal any important violations of the linearity assumption (section 2 of the Supplementary Material). When adjusted for SRI-5 response, as well as for age and sex, the odds of being in an active treatment group increased by a factor of about two for every 1–standard deviation increase in the LuMOS 2.0 score, with the adjusted association being somewhat stronger in the Illuminate-2 trial (adjusted OR [aOR] = 2.60, 95% CI: 2.08-3.26, $P < 0.0001$) than in Illuminate-1 data (aOR = 1.86, 95% CI: 1.54-2.25, $P < 0.0001$). On the other hand, in the Illuminate-1

Table 3. Comparisons of mean LuMOS scores across the randomization groups in the Illuminate-1 and Illuminate-2 trials

Trial	Group (n)	Mean LuMOS 2.0 score (95% CI)	Within-group SD	% SRI-5 responders (95% CI)
Illuminate-1	Placebo (n = 379)	−0.223 (−0.296 to −0.150)	0.726	29.3 (24.7 to 33.9)
Illuminate-1	TB-4weeks (n = 378)	+0.100 (+0.027 to +0.173)	0.725	35.2 (30.4 to 40.0)
Illuminate-1	TB-2weeks (n = 381)	+0.121 (+0.031 to +0.211)	0.898	31.8 (27.1 to 36.5)
Illuminate-2	Placebo (n = 376)	−0.235 (−0.310 to −0.160)	0.545	27.7 (23.2 to 32.2)
Illuminate-2	TB-4weeks (n = 376)	+0.103 (+0.033 to +0.173)	0.691	34.8 (30.0 to 39.6)
Illuminate-2	TB-2weeks (n = 372)	+0.207 (+0.133 to +0.281)	0.728	38.4 (33.5 to 43.3)

Abbreviations: CI, confidence interval; SD, standard deviation; SLE, systemic lupus erythematosus; SRI-5, SLE Responder Index-5; TB-2weeks, tabalumab 120 mg every 2 weeks; TB-4weeks, tabalumab 120 mg every 4 weeks.

Table 4. Effect sizes for discrimination between tabalumab groups and the placebo group in the Illuminate-1 and Illuminate-2 trials: LuMOS 2.0 versus SRI-5

Trial	Group (n)	LuMOS 2.0: effect size (95% CI)	SRI-5: effect size (95% CI)
Illuminate-1	TB-4weeks (n = 378)	0.44 (0.30 to 0.59)	0.13 (−0.02 to 0.27)
Illuminate-1	TB-2weeks (n = 381)	0.42 (0.27 to 0.56)	0.05 (−0.09 to 0.20)
Illuminate-2	TB-4weeks (n = 376)	0.54 (0.39 to 0.68)	0.15 (0.01 to 0.30)
Illuminate-2	TB-2weeks (n = 372)	0.69 (0.53 to 0.83)	0.23 (0.08 to 0.37)

Abbreviations: CI, confidence interval; SRI-5, Systemic Lupus Erythematosus Responder Index-5; TB-2weeks, tabalumab 120 mg every 2 weeks; TB-4weeks, tabalumab 120 mg every 4 weeks.

trial, SRI-5 had no association with the treatment, once adjusted for LuMOS 2.0 score (aOR for SRI-5 = 0.98; 95% CI: 0.75-1.27, P = 0.87). In Illuminate-2 analyses, the adjusted association between SRI-5 response and inclusion in a TB treatment group was only marginally significant (aOR = 1.33; 95% CI: 1.01-1.75, P = 0.044). Furthermore, in both trials, the adjusted associations were much stronger for LuMOS 2.0 scores than for SRI-5 responses (Wald statistics of 41.99 vs. 0.03, and 68.82 vs. 4.07 for Illuminate-1 and Illuminate-2, respectively). This pattern of results indicates that in both Illuminate trials, LuMOS 2.0 was able to identify important systematic differences between outcomes of patients treated with TB and outcomes of those randomized to a placebo that were not captured by the SRI-5. In contrast, once LuMOS 2.0 scores were taken into account, SRI-5 responses failed to improve discrimination between active treatment and placebo groups in Illuminate-1 participants and yielded only a marginal improvement in Illuminate-2 analyses.

DISCUSSION

A critical component of successful drug development in SLE is the availability of a simple rational outcome measure that is understandable to clinicians and relevant to their practices. This prompted us to develop a new evidence-based LuMOS (12). LuMOS was developed through multivariable analyses with the goal of optimizing discrimination between actively treated patients and those randomized to a placebo by aggregating information on longitudinal within-patient changes observed during the trial in a spectrum of relevant clinical and laboratory variables (12). Specifically, the multivariable LuMOS scoring formula was estimated based on the contrasts between the high-dose Bel treatment versus the placebo in the BLISS-76 trial, which originally established the effectiveness of this treatment (13) and, together with a similar BLISS-52 trial (14), led to US Food and Drug Administration (FDA) approval of Bel for SLE. In our reanalyses of both BLISS trials, LuMOS yielded highly statistically significant differences for each of the Bel groups compared with the placebo group, including those comparisons in which the SRI-4 used in the original analyses (13,14) failed to reach statistical significance. Moreover, for all comparisons, LuMOS indicated much stronger ES, as measured by Cohen's d statistic (15), than SRI-4 (12).

In the current article, we sought to validate LuMOS using unrelated lupus clinical trial data sets. To achieve this, we first refined the original LuMOS scoring formula to make it independent of assays used to measure the three laboratory biomarkers, which are now standardized using a z score transformation. We then conducted further external validation of the resulting standardized LuMOS 2.0 model using data from the Illuminate-1 and Illuminate-2 trials of TB (16,17) rather than Bel, which was used to develop LuMOS. In both Illuminate trials and for both TB treatment regimens tested in the trials, LuMOS 2.0 revealed statistically significant improvements in the outcomes, relative to patients randomized to the placebo. Thus, LuMOS 2.0 yielded consistent results for both Illuminate trials, in contrast to SRI-5, the primary outcome measure used in the original analyses of the same data, which showed significant improvements with TB only for the Illuminate-2 trial (17) but not for Illuminate-1 (16). Furthermore, for all comparisons between TB and the placebo in both trials, LuMOS 2.0 showed much stronger ES than SRI-5. The magnitude of these differences between the two outcome measures is underscored by the fact that even the lower bounds of 95% CIs for LuMOS-based ES were systematically higher than the upper bounds of the corresponding CIs for SRI-5.

Additional exploratory multivariable analyses yielded further insights regarding the comparison of responsiveness to change of LuMOS 2.0 and SRI-5. The results demonstrated that LuMOS 2.0 scores provided a statistically significant clear discrimination between changes observed during either Illuminate trial in actively treated participants and changes observed in those on a placebo, even after adjustment for SRI-5. Thus, even among the trial participants with the same SRI-5 responses, those with higher LuMOS 2.0 scores are significantly more likely to be treated with TB, with the odds approximately doubling for every 1–standard deviation increase in LuMOS 2.0 score. In contrast, among participants with the same LuMOS 2.0 score, the differences between SRI-5 responses in the TB groups and those in the placebo group were only marginally significant (P = 0.044) in Illuminate-2 data (17) and convincingly nonsignificant for the Illuminate-1 trial (P = 0.87) (16). Overall, these results suggest that LuMOS is able to identify some systematic differences in the pattern of changes in relevant clinical and laboratory variables observed during the trial in actively treated patients, relative to spontaneous changes that occur in the placebo group, that are not captured by SRI-5, while also accounting for practically all meaningful differences that are reflected by SRI-5.

Two formal properties of LuMOS may partly explain its better ability to identify trial participants who receive an active treatment

rather than a placebo. Firstly, a continuous outcome measure, such as the LuMOS 2.0 score, is in general expected to offer better precision and higher statistical power than a binary responder index, such as SRI-4 or SRI-5. Secondly, the LuMOS scoring formula involves multiplying changes observed during the trial in particular variables by the corresponding adjusted regression coefficients from multivariable analyses (12). Thus, the final LuMOS score reflects both the magnitude of the changes in the relevant laboratory and clinical variables and their evidence-based numerical relative importance weights, resulting in a more refined and informative aggregation of the multivariable pattern of changes than just counting the number of variables for which the observed change exceeds a predetermined threshold.

Our current work aims to refine outcome measurement in SLE trials in which the main focus is on comparing the average responses of actively treated participants versus those randomized to a placebo (ie, on contrasting the mean LuMOS 2.0 values in the corresponding groups). Further research may explore if and how LuMOS 2.0 scores may be used to assess longitudinal within-patient changes in SLE disease activity of individual patients.

Some important limitations of both the LuMOS approach and the analyses reported in this study have to be recognized. As we acknowledged earlier, the fundamental, likely unsurmountable, difficulty is related to an absence of an independent gold standard for a clinically relevant and empirically measurable global improvement in SLE disease manifestations (12). Indeed, empirical comparisons of the independent assessments of real patient-derived clinical vignettes by several SLE experts revealed important differences in opinions regarding whether individual patients did or did not show a clinically meaningful improvement, even when the experts were given exactly the same information (5). In this situation, when developing LuMOS, we had to rely on discrimination between outcomes observed in patients treated with an active FDA-approved SLE treatment (13), such as high-dose Bel, and outcomes observed in patients who received a placebo. Whereas it is likely that, on average, this contrast is a reasonable proxy for the differences associated with a hypothetical gold standard of a clinically relevant improvement, it is clear that some Bel-treated patients had no improvements during the trial. Thus, the dependent variable in our multivariable analyses through which LuMOS was developed may be affected by some misclassification, likely inducing both bias toward the null in the estimated associations and some, less predictable, inaccuracy in the resulting numerical weights assigned to changes in the particular component variables. From this perspective, we find it encouraging that, in spite of these limitations, LuMOS 2.0 performed very well in discriminating between actively treated patients and patients on a placebo in the current analyses of two independent trials of a different drug from Bel used to develop the LuMOS scoring formulas.

On the other hand, it is important to take into account that TB evaluated in the two Illuminate trials (16,17) on which our current validation analyses rely has a similar action as Bel assessed in the BLISS-76 trial (13), the results of which were used to develop LuMOS and LuMOS 2.0. As a consequence, if, for example, the relatively high weight assigned in LuMOS to changes in the C4 component (Table 2) reflects a strong impact of autoantibody secretion by B cell lineage cells, common to both drugs (13,14,16,17), this might have favored LuMOS in our validation analyses. Therefore, a priority for our future research is to attempt further validation of LuMOS 2.0 using data from successful SLE trials that assessed other effective treatment(s) with different molecular targets than Bel and TB. Ideally, such new trials could also provide measures of some other relevant variables, such as patient-reported outcomes, which were not systematically collected in the BLISS-52 and BLISS-76 and Illuminate-1 and Illuminate-2 trials (13,14,16,17) but are likely to yield independent, clinically important information relevant to identify responders to an effective treatment. Access to such additional data sources would allow us to reassess the current LuMOS and LuMOS 2.0 models through new multivariable analyses based on random forests (25) and other promising, recently developed variable selection methods (26,27).

In summary, our validation analyses suggest that a versatile standardized LuMOS 2.0 model may help identify effective new SLE treatments by identifying responders more accurately. Yet further validation using data from successful trials of SLE drugs with different biochemical properties from Bel and TB is required before LuMOS can be established as a generic SLE responder index.

ACKNOWLEDGMENTS

Michal Abrahamowicz is a James McGill Professor of Biostatistics at McGill University. The authors thank Dr. Marie-Eve Beauchamp and Mrs. Sofia Bamboulas for their assistance with the preparation of the manuscript.

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. M. Abrahamowicz 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. M. Abrahamowicz, Lipsky.

Acquisition of data. Lipsky.

Analysis and interpretation of data. M. Abrahamowicz, M. I. Abrahamowicz, Lipsky.

REFERENCES

1. Aringer M, Strand V. Endpoints for randomised controlled trials in systemic lupus erythematosus. Clin Exp Rheumatol 2012;30:147–51.
2. Mahieu MA, Strand V, Simon LS, Lipsky PE, Ramsey-Goldman R. A critical review of clinical trials in systemic lupus erythematosus. Lupus 2016;25:1122–40.
3. Forbess LJ, Bresee C, Wallace DJ, Weisman MH. Failure of a systemic lupus erythematosus response index developed from

clinical trial data: lessons examined and learned. Lupus 2017;26: 909–16.

4. Liang MH, Corzillius M, Bae SC, Fortin P, Esdaile JM, Abrahamowicz M. A conceptual framework for clinical trials in SLE and other multisystem diseases. Lupus 1999;8:570–80.
5. American College of Rheumatology Ad Hoc Committee on Systemic Lupus Erythematosus Response Criteria. The American College of Rheumatology response criteria for systemic lupus erythematosus clinical trials: measures of overall disease activity. Arthritis Rheum 2004;50:3418–26.
6. Bombardier C, Gladman DD, Urowitz MB, Caron D, Chang DH, and the Committee on Prognosis Studies in SLE. Derivation of the SLEDAI: a disease activity index for lupus patients. Arthritis Rheum 1992;35: 630–40.
7. 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.
8. Liang MH, Socher SA, Larson MG, Schur PH. Reliability and validity of six systems for the clinical assessment of disease activity in systemic lupus erythematosus. Arthritis Rheum 1989;32:1107–18.
9. Bae SC, Koh HK, Chang DK, Kim MH, Park JK, Kim SY. Reliability and validity of systemic lupus erythematosus measure-revised (SLAM-R) for measuring clinical disease activity in systemic lupus erythematosus. Lupus 2001;10:405–9.
10. Petri M, Kim MY, Kalunian KC, Grossman J, Hahn BH, Sammaritano LR, et al. Combined oral contraceptives in women with systemic lupus erythematosus. N Engl J Med 2005;353:2550–8.
11. Furie RA, Petri MA, Wallace DJ, Ginzler EM, Merrill JT, Stohl W, et al. Novel evidence-based systemic lupus erythematosus responder index. Arthritis Rheum 2009;61:1143–51.
12. Abrahamowicz M, Esdaile JM, Ramsey-Goldman R, Simon LS, Strand V, Lipsky PE. Development and validation of a novel evidence-based lupus multivariable outcome score for clinical trials. Arthritis Rheum 2018;70:1450–8.
13. Furie R, Petri M, Zamani O, Cervera R, Wallace DJ, Tegzova D, 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.
14. Navarra SV, Guzman RM, Gallacher AE, Hall S, Levy RA, Jimenez RE, et al. Belimumab, a BLyS-specific inhibitor, reduced disease activity, flares and prednisone use in patients with active SLE: efficacy and safety results from the phase 3 BLISS-52 study. Lancet 2011;377: 721–31.
15. Cohen J. Statistical power analyses for the behavioral sciences. New York: Academic Press; 1969.
16. Isenberg DA, Petri M, Kalunian K, Tanaka Y, Urowitz MB, Hoffman RW, 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.
17. 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.
18. Huberty CJ. A history of effect size indices. Educ Psychol Meas 2002; 62:227–40.
19. Hedges L, Olkin I. Statistical methods for meta-analysis. New York: Academic Press; 1985.
20. Hastie TJ, Tibshirani RJ. Generalized additive models. New York: Chapman & Hall/CRC; 1990.
21. Greenland S. Dose-response and trend analysis in epidemiology: alternatives to categorical analysis. Epidemiology 1995;6:356–65.
22. Abrahamowicz M, du Berger R, Grover SA. Flexible modeling of the effects of serum cholesterol on coronary heart disease mortality. Am J Epidemiol 1997;145:714–29.
23. Akaike H. A new look at the statistical model identification. IEEE Trans Automat Contr 1974;19:716–23.
24. Abrahamowicz M, Beauchamp ME, Sylvestre MP. Comparison of alternative models for linking drug exposure with adverse effects. Stat Med 2012;31:1014–30.
25. Wright MN, Dankowski T, Ziegler A. Unbiased split variable selection for random survival forests using maximally selected rank statistics. Stat Med 2017;36:1272–84.
26. Dunkler D, Plischke M, Leffondre K, Heinze G. Augmented backward elimination: a pragmatic and purposeful way to develop statistical models. PLoS One 2014;9:e113677.
27. Sauerbrei W, Perperoglou A, Schmid M, Abrahamowicz M, Becher H, Blinder H, et al. State-of-the-art in selection of variables and functional forms in multivariable analysis: outstanding issues. Diagn Progn Res 2020;4:3.