



ScienceDirect

Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/jval

Preference-Based Assessments

Valuing Health Gain from Composite Response Endpoints for Multisystem Diseases



Sean P. Gavan, PhD, Ian N. Bruce, MD, Katherine Payne, PhD, on behalf of the MASTERPLANS Consortium

ABSTRACT

Objectives: This study aimed to demonstrate how to estimate the value of health gain after patients with a multisystem disease achieve a condition-specific composite response endpoint.

Methods: Data from patients treated in routine practice with an exemplar multisystem disease (systemic lupus erythematosus) were extracted from a national register (British Isles Lupus Assessment Group Biologics Register). Two bespoke composite response endpoints (Major Clinical Response and Improvement) were developed in advance of this study. Difference-in-differences regression compared health utility values (3-level version of EQ-5D; UK tariff) over 6 months for responders and nonresponders. Bootstrapped regression estimated the incremental quality-adjusted life-years (QALYs), probability of QALY gain after achieving the response criteria, and population monetary benefit of response.

Results: Within the sample ($n = 171$), 18.2% achieved Major Clinical Response and 49.1% achieved Improvement at 6 months. Incremental health utility values were 0.0923 for Major Clinical Response and 0.0454 for Improvement. Expected incremental QALY gain at 6 months was 0.020 for Major Clinical Response and 0.012 for Improvement. Probability of QALY gain after achieving the response criteria was 77.6% for Major Clinical Response and 72.7% for Improvement. Population monetary benefit of response was £1106458 for Major Clinical Response and £649134 for Improvement.

Conclusions: Bespoke composite response endpoints are becoming more common to measure treatment response for multisystem diseases in trials and observational studies. Health technology assessment agencies face a growing challenge to establish whether these endpoints correspond with improved health gain. Health utility values can generate this evidence to enhance the usefulness of composite response endpoints for health technology assessment, decision making, and economic evaluation.

Keywords: composite response endpoint, health state utility, monetary benefit, multisystem disease, systemic lupus erythematosus.

VALUE HEALTH. 2023; 26(1):115–122

Introduction

Condition-specific composite response endpoints are being developed to improve the ability to determine whether patients with multisystem diseases benefit from treatment within clinical trials and observational studies,^{1,2} but health technology assessment (HTA) agencies and decision makers face a challenge to ascertain whether these endpoints correspond with improved health gain.^{3,4} This challenge can be resolved by demonstrating a clear link between achieving a composite response endpoint and the likely health gain expressed in terms of an HTA agency's preferred instrument to measure health state utility.^{5,6} Multidimensional health-related quality of life instruments with tariffs informed by population preferences are an underused resource to value the health gain from composite response endpoints.⁷ Evidence of this health gain is vital information to justify a composite response endpoint's usefulness for decision making

and should be considered alongside its clinical relevance and statistical properties.

HTA agencies are experiencing a rise in treatments for multisystem diseases, which are characterized by heterogeneous activity across more than one organ system.^{8,9} Composite response endpoints refer to explicit criteria, which incorporate 2 or more independent outcomes into a single dichotomous outcome to determine whether patients respond to treatment over time ("responder" or "nonresponder").⁴ These composite response endpoints are thought to better capture how multisystem diseases improve after effective treatment.¹⁰ By contrast, response endpoints defined by a single instrument may not be sensitive to change if they do not measure all important improvements in disease activity across different organ systems.¹¹

One common justification for developing a bespoke composite response endpoint is that improvement may be characterized best by reference to multiple outcomes.¹² By combining these

outcomes into a single index, the need for a larger sample size to achieve statistical power is reduced.¹² Nevertheless, cost-effectiveness analyses are performed within a Bayesian decision analysis framework,¹³ and accordingly, corresponding changes in health state utility can be evaluated in terms of their distribution, expected value, and probability that patients categorized as “responders” will experience health gain. The EQ-5D (3L and 5L versions), with accompanying population preference weights, is often favored by decision makers internationally to estimate health state utility values and calculate quality-adjusted life-years (QALYs).⁵ Therefore, the value of health gain from a composite response endpoint can be quantified by outcomes such as the expected incremental improvements in EQ-5D utility score, QALYs, and, equivalently, the expected monetary benefit of responding to treatment. Expressing health gain in terms of outcomes preferred by HTA agencies, instead of condition-specific outcomes alone, will justify the importance of a composite response endpoint for decision making. Failing to demonstrate this evidence will ultimately reduce the usefulness of pivotal studies that use a bespoke composite response endpoint to estimate treatment efficacy or effectiveness.

The multisystem disease systemic lupus erythematosus (SLE) provides a good exemplar to illustrate the value of expressing the health gain associated with bespoke disease-specific composite response endpoints in terms of outcomes favored by decision makers.¹⁴ Clinical trials in SLE have remained a challenge with only 2 treatments achieving regulatory approval in the past 60 years.^{10,15} In 2005, the US Food and Drug Administration produced initial guidance (finalized in 2010), which recommended composite response endpoints within trials for SLE to capture whether outcomes are nondeteriorating across more than one domain within the same measure.¹⁶ Two recent examples of composite response endpoints created for SLE, informed by this guidance, are the British Isles Lupus Assessment Group-based Composite Lupus Assessment (BICLA), first used in the phase IIb study of epratuzumab,¹⁷ and the Systemic Lupus Erythematosus Responder Index 4 (SRI-4), developed from the phase II trial data for belimumab.¹⁸ The BICLA and the SRI-4 both incorporate observations from 3 separate condition-specific instruments that measure disease activity (the British Isles Lupus Assessment Group [BILAG] 2004 Index,¹⁹ the Systemic Lupus Erythematosus Disease Activity Index 2000 [SLEDAI-2K],²⁰ and the Physician Global Assessment²¹) to determine whether patients are classified dichotomously as either a “responder” or a “nonresponder.” These composite response endpoints have certain limitations. For

example, the BICLA is less sensitive for patients with high baseline organ involvement and the SRI-4 does not detect partial improvement in an organ system.¹¹ Nevertheless, they have enhanced the quantification of disease improvement across a patient population with heterogeneous clinical manifestations and contributed to the recent success of trials in SLE.¹⁰ Therefore, defining treatment outcomes according to composite response endpoints is likely to remain the standard for the foreseeable future. HTA agencies are now confronted with the challenge of comparing evidence from studies that use different composite response endpoints and evaluating the importance of these endpoints according to their preferred instrument to value health gain. Therefore, to address this growing challenge across diseases that affect more than one organ system, this study aims to demonstrate how to estimate the value of health gain after patients with a multisystem disease achieve a condition-specific composite response endpoint.

Methods

This study uses regression analyses of data collected as part of routine (nontrial) practice within a national patient register. A case study is presented to illustrate how health state utility values (from 3-level version of EQ-5D [EQ-5D-3L] profiles informed by general population preference weights) were used to value the health gain associated with 2 bespoke composite response endpoints for a multisystem disease.

Case Study

SLE is an autoimmune disease with heterogeneous symptoms among patients that is caused by disease activity across different organ systems. Globally, the incidence of SLE ranges between 1.5 and 11 per 100 000 person-years.²² SLE has a 2 to 3 times higher mortality risk than the general population.²² Treatment with antimalarials, immunosuppressants, biologic agents, and glucocorticoids is intended to reduce active disease and prevent organ damage.²³ The “Maximizing SLE Therapeutic Potential by Application of Novel and Stratified Approaches” program was a Medical Research Council Precision Medicine Consortium that aimed to develop statistical algorithms that could predict the likelihood of response to treatments used routinely to manage SLE.²⁴ As part of this program, 2 bespoke condition-specific composite response endpoints for SLE were developed a priori to define criteria for achieving “Major Clinical Response” or “Improvement” after

Table 1. Criteria for achieving both composite response endpoints.

Criterion	Improvement	Major Clinical Response
BILAG 2004 index	<ul style="list-style-type: none"> No BILAG A score and ≤ 1 BILAG B score at 6 months 	<ul style="list-style-type: none"> No BILAG A score and no BILAG B score at 6 months
SLEDAI-2K	If baseline SLEDAI-2K ≤ 4 : <ul style="list-style-type: none"> SLEDAI-2K ≤ 4 at 6 months If baseline SLEDAI-2K > 4 : <ul style="list-style-type: none"> Any reduction in SLEDAI-2K at 6 months 	<ul style="list-style-type: none"> SLEDAI ≤ 4 at 6 months
Corticosteroid dose	If baseline dose ≤ 10 mg/day: <ul style="list-style-type: none"> Dose ≤ 10 mg/day at 6 months If baseline dose ≥ 11 mg/day but < 20 mg/day: <ul style="list-style-type: none"> Dose ≤ 10 mg/day at 6 months If baseline dose ≥ 20 mg/day: <ul style="list-style-type: none"> Dose ≤ 15 mg/day at 6 months 	<ul style="list-style-type: none"> Dose ≤ 7.5 mg/day at 6 months

BILAG indicates British Isles Lupus Assessment Group; SLEDAI-2K, Systemic Lupus Erythematosus Disease Activity Index 2000.

Table 2. Characteristics of the estimation sample (n = 171).

Variable	Value
Age, years (SD)	43.55 (13.98)
Female	89.5%
Pretreatment disease activity	
SLEDAI-2K (SD)	8.51 (5.79)
At least 1 BILAG A score	60.8%
At least 2 BILAG B scores	48.5%
Treatment received	
Rituximab	85.4%
Mycophenolate mofetil	9.4%
Belimumab	5.3%
Endpoint achieved at 6 months	
Improvement	49.1%
Major Clinical Response	18.1%

Note. Criteria for Major Clinical Response and Improvement are presented in Table 1.

BILAG indicates British Isles Lupus Assessment Group; SLEDAI-2K, Systemic Lupus Erythematosus Disease Activity Index 2000.

treatment. Patients who are prescribed a biologic treatment for SLE in England are required to enroll with a national patient register as a condition of reimbursement.^{25,26} This national register collects disease-specific outcomes and health-related quality of life (EQ-5D-3L) data. The routine collection of these data in parallel with evidence of response to treatments prescribed in a clinical setting provided an opportunity to estimate the value of the health gain associated with the 2 bespoke composite response endpoints.

Data Source

The data for this study were obtained from the British Isles Lupus Assessment Group Biologics Register (BILAG-BR).²⁷ The BILAG-BR records clinical and demographic data from all patients in England who received a biologic treatment for SLE (belimumab, rituximab) and from some patients who received nonbiologic immunosuppressant treatment (mycophenolate mofetil), at baseline (before receiving treatment) and defined follow-up periods. Data were collected within routine clinical practice from 59 specialist centers across the UK between 2010 and 2019.

Instruments Measured

Two disease-specific instruments measured disease activity at baseline and 6 months. The BILAG-2004 Index measured SLE-specific disease activity across 9 organ systems.¹⁹ Activity in each system was scored on an ordinal scale of A to E, where A means severe active disease requiring high-dose systemic treatment, B means moderate disease activity requiring low-dose corticosteroid or antimalarial treatment, C means mild disease activity, D means inactive disease that was active previously, and E means the system has never been active previously. The SLEDAI-2K measured global disease activity in 9 organ systems across 24 items, weighted by severity, and has a score between 0 and 105.²⁰ The follow-up duration reflected the timing of treatment continuation decision making within clinical practice, which is informed by disease improvement at 6 months.^{25,26} Health-related quality of life was recorded at baseline and 6 months with the EQ-5D-3L.²⁸ The EQ-5D-3L has 5 domains (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) and 3 levels (no problems, some problems, and extreme problems) to describe 243 unique health states. The EQ-5D-3L tariff for

the general population in the United Kingdom was used to estimate health state utility values on a scale of 1 (perfect health) to 0 (dead), where states worse than dead (< 0) were possible.²⁹ Patients' demographics (age and female sex) and initial treatment were recorded at baseline. Oral corticosteroid dose (mg) was recorded at baseline and 6 months.

Composite Response Endpoint Criteria

Two bespoke composite response endpoints (Major Clinical Response and Improvement) were developed by clinicians and patient experts who were part of the Maximizing SLE Therapeutic Potential by Application of Novel and Stratified Approaches Consortium. The criteria of these 2 endpoints comprised 3 items: the change in each patient's (1) BILAG score, (2) SLEDAI-2K score, and (3) corticosteroid dose between baseline and 6 months (Table 1). For each endpoint, patients who met these criteria were classified dichotomously as a "responder" and "nonresponder" otherwise. The criteria for Major Clinical Response were stricter than those for Improvement and, subsequently, was hypothesized to characterize a greater health gain.

Estimation Sample

The estimation sample comprised patients with (1) high baseline disease activity before treatment (at least one BILAG A score, 2 BILAG B scores, or SLEDAI-2K \geq 6), (2) sufficient data to determine whether both composite response endpoints were achieved at 6 months, and (3) complete EQ-5D-3L observations at baseline and 6 months.

Analysis

Descriptive statistics reported the baseline characteristics of the estimation sample according to their average demographics, treatment received in routine practice, and disease activity (mean SLEDAI-2K and percentage of patients with at least 1 BILAG A or 2 BILAG B scores). The percentage of patients who fulfilled the Improvement or Major Clinical Response endpoint criteria at 6 months were also reported.

For each composite response endpoint, the sample was then stratified by those who achieved the endpoint ("responder") and those who did not ("nonresponder"). To understand how health state utility values changed after achieving each endpoint, the mean EQ-5D-3L utility scores at baseline and 6 months and the mean difference (between baseline and 6 months) were reported separately for patients in the "responder" and "nonresponder"

Table 3. Mean health state utility values for both composite response endpoints.

Time point	Health state utility	
	Responder	Nonresponder
Improvement at 6 months (n = 171)		
Baseline	0.449 (0.352)	0.520 (0.307)
6 months	0.499 (0.352)	0.525 (0.344)
Difference	0.050 (0.287)	0.004 (0.277)
Major Clinical Response at 6 months (n = 171)		
Baseline	0.344 (0.371)	0.517 (0.314)
6 -months	0.446 (0.364)	0.527 (0.343)
Difference	0.102 (0.289)	0.010 (0.279)

Note. Health state utility measured by EQ-5D-3L and valued with the UK population tariff. Standard deviation presented in parentheses. Criteria for Major Clinical Response and Improvement are presented in Table 1

Table 4. Regression results for both composite response endpoints.

Variable	Improvement	Major Clinical Response
R	-0.0712 (0.0519)	-0.1726* (0.0667)
T	0.0044 (0.0514)	0.0100 (0.0402)
R × T	0.0454 (0.0734)	0.0923 (0.0944)
Intercept	0.5203 [†] (0.0364)	0.5166 [†] (0.0284)
N	342	342
r ²	0.0078	0.0250

Note. Standard errors presented in parentheses; Criteria for Major Clinical Response and Improvement are presented in Table 1.

N indicates the sample size; R, a binary variable equal to one if the patient achieved the specific composite response endpoint and zero otherwise; r², the coefficient of determination; R × T, an interaction term; T, a binary variable equal to one if the observation was measured at 6 months and zero if at baseline.

*Statistical significance at 1%.

[†]Statistical significance at 5%.

subgroups. Two-period difference-in-differences ordinary least squares regressions on the pooled baseline and 6-month observations estimated the incremental change in utility score associated with achieving each composite response endpoint (relative to not achieving the endpoint).³⁰ The functional form of each regression is reported in (Eq. 1), where Y is the EQ-5D-3L utility score, R is a binary variable equal to one if the patient achieved the specific composite response endpoint and zero otherwise, T is a binary variable equal to one if the observation was measured at 6 months and zero if at baseline, and (RT) is an interaction term.

$$Y = \beta_0 + \beta_1 R + \beta_2 T + \beta_3 RT \quad (1)$$

The estimated parameter for the interaction term (β_3) is most relevant for this study because it is equivalent to the incremental change in health state utility for patients who achieve the endpoint (the difference in EQ-5D-3L utility at 6 months for responders less the difference in EQ-5D-3L utility at 6 months for nonresponders). The remaining estimated parameters are interpreted as follows: β_0 is the mean EQ-5D-3L utility for nonresponders at baseline, $\beta_0 + \beta_1$ is the mean EQ-5D-3L utility for responders at baseline, and β_2 is the estimated difference in the EQ-5D-3L utility between baseline and 6 months for nonresponders.

QALYs were calculated by the trapezium rule for each patient.³¹ For each endpoint, the incremental QALY gain at 6 months for responders versus nonresponders was estimated by ordinary least squares regression, which controlled for baseline utility scores with 10 000 bootstrap replications.³¹ The density function of the incremental QALY gain in each replication was plot separately for each composite response endpoint (kernel density estimator) to facilitate a comparison between the distributions in terms of expected values and parameter uncertainty. The probability that patients experience a positive QALY gain after achieving each composite response endpoint was estimated by the proportion of replications that were greater than zero. The incremental health gain was converted to monetary units (monetary benefit of response) by multiplying the expected incremental QALY gain by the threshold value used to determine cost-effectiveness (k).³² A cost-effectiveness threshold value of £20 000 per QALY gained was used because this is the lower value of the range used by the National Institute for Health and Care Excellence in their health technology appraisal process.³³ Per patient estimates of incremental QALY gain (ΔQ) and monetary benefit were multiplied by

the number of incident patients (I) over 1 year to provide population-level estimates. The population monetary benefit of response ($PMBR_p$) was plot separately for Improvement and Major Clinical Response over a range of values that corresponded with the proportion of patients in incident population who achieved each endpoint (Eq. 2). All analyses were performed in Stata version 16.1 (StataCorp LLC, College Station, TX).³⁴

$$PMBR_p = \Delta Q k I p \quad (2)$$

Ethics

Ethical approval for the BILAG-BR was obtained from the National Research Ethics Service (NRES) Committee North West-Greater Manchester West (REC: 09/H1014/64) and the local research and development departments at participating sites. All patients provided a written informed consent at the time of registration to the study.

Results

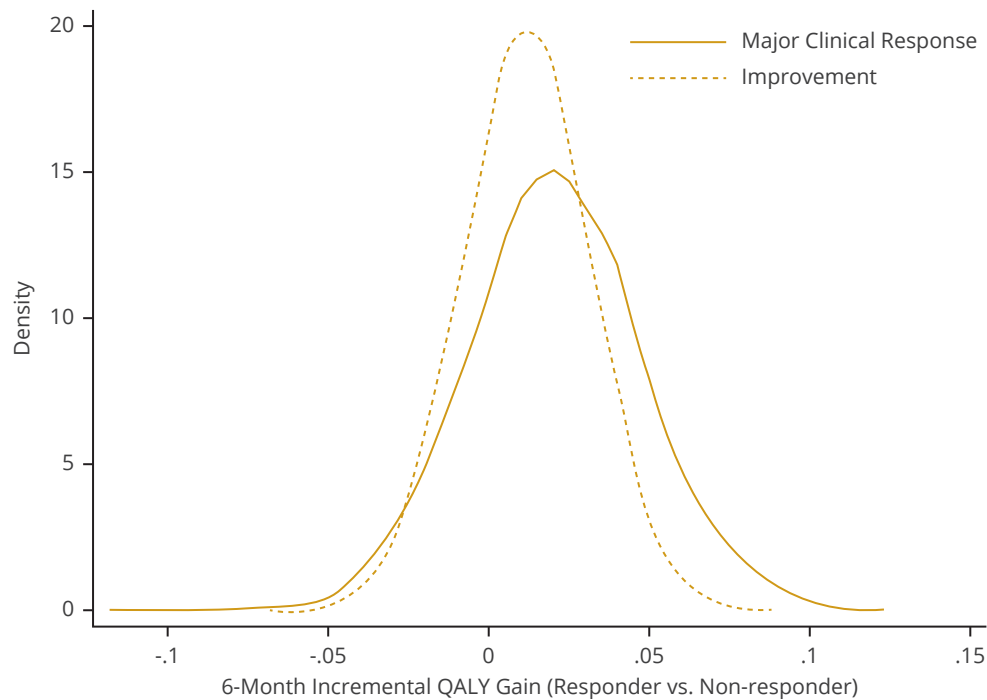
There were 776 individuals in the register with available data to determine whether the composite response endpoints were achieved at 6 months. Of this data set, the estimation sample comprised the 171 patients with complete EQ-5D-3L observations at baseline and 6 months. The baseline (pretreatment) descriptive statistics of the estimation sample and the percentage who achieved the Improvement and Major Clinical Response endpoints are presented in Table 2. On average, the sample had severe active disease (mean SLEDAI-2K: 8.51); more than half of patients had at least one BILAG A score and 48.5% of patients had at least 2 BILAG B scores before receiving treatment. Most patients received a biologic treatment at baseline (rituximab, 85.4%; belimumab, 5.3%). At 6 months, 49.1% of patients achieved the Improvement endpoint and 18.1% of patients achieved the Major Clinical Response endpoint.

Table 3 reports separately for each composite response endpoint, the mean health state utility values at baseline and 6 months, and the difference over time, stratified by whether patients achieved each endpoint (responder) or not (nonresponder). For both composite response endpoints, patients who achieved the response criteria had, on average, a lower baseline health state utility value than patients who did not achieve the response criteria (Improvement, 0.449 vs 0.520; Major Clinical Response, 0.344 vs 0.517). The mean health state utility values increased over 6 months irrespective of whether patients met either endpoint. The magnitude of this increase was larger for patients who achieved each endpoint. The increase in mean health state utility values for patients who achieved the Major Clinical Response endpoint (difference, 0.102) was approximately 2 times greater than that for patients who achieved the Improvement endpoint (difference, 0.050).

The estimated coefficients from the difference-in-differences regressions separately for each composite response endpoint are presented in Table 4. The positive coefficient on the interaction term indicates that, for both endpoints, the incremental gain in health state utility was positive for patients who achieved the respective endpoint compared with those who did not. The estimated incremental gain in health state utility for patients who achieved the Major Clinical Response endpoint ($\beta_3 = 0.0923$) at 6 months was larger than that for the Improvement endpoint ($\beta_3 = 0.0454$).

The 6-month per patient incremental QALY gain for response, after controlling for baseline health state utility, was estimated as 0.020 QALYs for Major Clinical Response and 0.012 QALYs for

Figure 1. Density function of bootstrapped incremental QALY gain by composite response endpoint.



QALY indicates quality-adjusted life-year. Criteria for Major Clinical Response and Improvement are presented in Table 1. Kernel density estimator of 10 000 bootstrap replications.

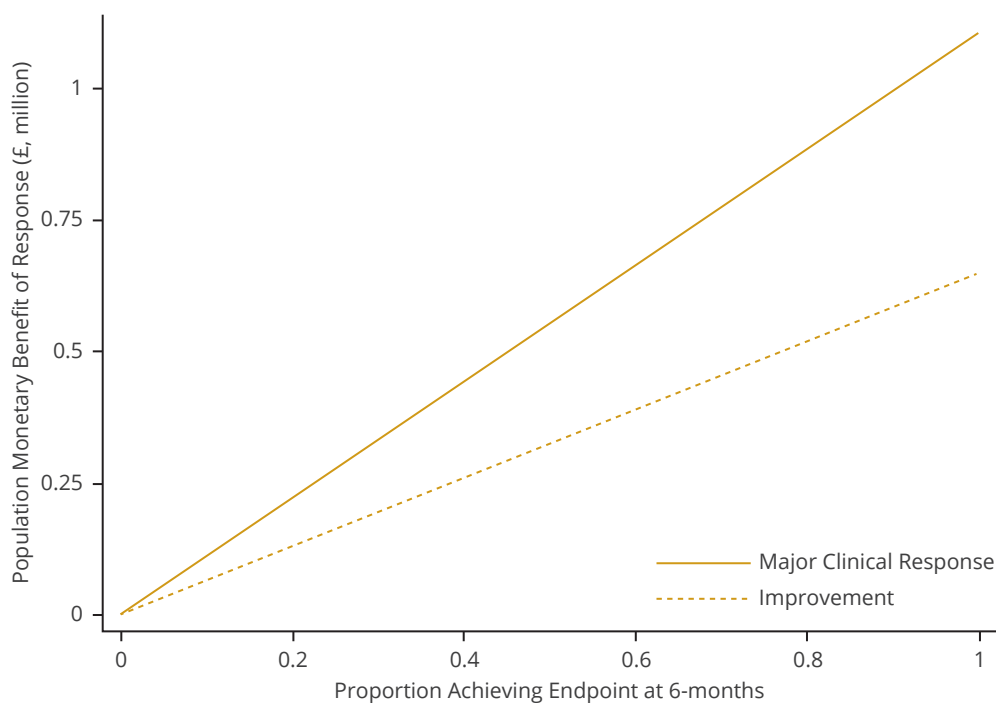
Improvement. The 6-month per patient incremental monetary benefit of response was £398 for Major Clinical Response and £233 for Improvement. The density function of the simulated incremental QALY gains from achieving each composite response endpoint (relative to being a nonresponder) is illustrated in Figure 1. The sampled distribution of the incremental QALY gain was wider for the Major Clinical Response endpoint than for the Improvement endpoint, which showed that parameter uncertainty associated with the true value of the incremental QALY gain was greater for patients who achieved Major Clinical Response. The distribution of the QALY gain for patients who achieved Major Clinical Response was further to the right than for patients who achieved Improvement, which indicated that Major Clinical Response corresponded with a higher expected incremental QALY gain. By reference to the proportion of simulations that were greater than zero, the probability that achieving each endpoint was associated with a positive QALY gain was 77.6% for Major Clinical Response and 72.7% for Improvement.

The estimated number of new SLE cases per year was 2776, based on Rees et al's³⁵ estimated incidence of 4.91 per 100 000 using national electronic healthcare record data in the UK (Clinical Practice Research Datalink) and assuming a general population of 56.55 million individuals.³⁶ Figure 2 illustrates how the 6-month population monetary benefit of response for this incident population changes as the proportion of patients who achieve each endpoint increases. If all patients achieve Major Clinical Response, there will be a population benefit of response equivalent to 55.32 QALYs or £1 106 458. Similarly, if all patients achieve Improvement, there will be a population benefit of response equivalent to 32.46 QALYs or £649 134. Based on the proportion of patients who currently achieve each endpoint in routine practice (see Table 2), the expected population monetary benefit of response for Major Clinical Response is £200 600 (or 10.03 QALYs) and for Improvement is £318 854 (or 15.94 QALYs).

Discussion

This study demonstrated how bespoke disease-specific composite response endpoints can be strengthened by linking them with improvements in health gain according to outcomes that are relevant to HTA agencies. In the case study, patients who achieved the Improvement and Major Clinical Response endpoints at 6 months had higher increases in health state utility, QALYs, and monetary benefit of response than patients who did not achieve these endpoints. Valuing the health gain of a composite response endpoint can guide the criteria for meeting these endpoints as they are developed, help inform trial design by selecting the most valuable composite response endpoint from a range of competing alternatives, and improve the usefulness of future research studies that use the endpoint within the context of informing HTA and decision making.

The need to transform disease-specific instruments into health state utility values to support decision making has long been recognized in the mapping literature.³⁷ Nevertheless, composite response endpoints measure the change of disease activity over time after treatment (a measure of effectiveness), whereas mapping studies estimate a cross-sectional association between a disease-specific instrument and health state utility in the same time period. To value the health gain associated with a composite response endpoint, analysts will first require patient-level baseline and follow-up data. At each time period, these data should include observations of health state utility and an indicator variable for whether the composite response endpoint was met at the defined follow-up. Regression analyses to control for baseline utility with nonparametric bootstrapping, equivalent to the approach for within-trial economic evaluations,³¹ can then estimate the incremental health gain of meeting the endpoint compared with not responding. Large observational studies (patient registers, cohort studies) and clinical trials will

Figure 2. Population monetary benefit of response by composite response endpoint.

QALY indicates quality-adjusted life-year. Criteria for Major Clinical Response and Improvement are presented in Table 1. Population monetary benefit of response (£, million) calculated assuming an incident population of 2776 patients and a cost-effectiveness threshold of £20 000 per QALY gained.

both provide relevant patient-level data to complete these analyses.

Valuing health gain according to changes in health state utility values or QALYs are typically an essential input for cost-effectiveness analyses. A recent review of economic evaluations in SLE found estimates of utility values were obtained from small samples or were reported incompletely.³⁸ The results from the present study build on this finding and provide a set of health state utility values associated with achieving the Improvement and Major Clinical Response endpoints, which can, subsequently, be used as a source of evidence for input parameters in future cost-effectiveness analyses. Parameter uncertainty in these utility values can be propagated appropriately within a probabilistic analysis by reference to the reported standard errors. In turn, the evidence reported by this study can help to improve the robustness of future analyses, which aim to estimate the relative cost-effectiveness of new and existing treatment strategies for SLE.

Previous studies have compared responders and non-responders with existing composite response endpoints for SLE with other health-related quality of life instruments. Furie et al³⁹ estimated that a greater percentage of patients who achieved the BICLA composite response endpoint at 52 weeks, than non-responding patients, experienced at least a 3-point improvement in the Short-Form 36 (SF-36) Physical Component Summary score (57.9% vs 12.8%) and at least a 4.6-point improvement in the SF-36 Mental Component Summary score (42.6% vs 12.3%) as part of a post hoc analysis of the pooled anifrolumab phase III trial data. Similarly, Furie et al⁴⁰ used pooled 52-week observations from the belimumab in subjects with systemic lupus erythematosus (BLISS)-52 and BLISS-76 trial data for belimumab and found that patients who achieved the SRI composite response endpoint had a greater improvement in the SF-36 Physical Component Summary (4.9 vs 2.6) and SF-36 Mental Component Summary (4.4 vs 1.7)

than nonresponding patients. If there is a positive correlation between the SF-36 and EQ-5D-3L in people with SLE,⁴¹ these findings could suggest that existing endpoints may also improve EQ-5D scores. Direct comparative evidence of how existing composite response endpoints perform according to instruments preferred by HTA agencies will be valuable evidence to support future decision making.

One limitation of this study was that the composite response endpoints were measured at 6 months only. The duration of treatment outcomes for SLE varies among individuals, and retreatment can be effective for patients who relapse.⁴² Beyond a 6-month time horizon, the health gain attributable to the composite response endpoints will be determined by the proportion who sustain their 6-month outcomes and effectiveness of retreatment in relapsing cases. A second limitation was that fewer patients achieved the Major Clinical Response endpoint (18.1%) than the Improvement endpoint (49.1%), which may increase parameter uncertainty around the expected QALY gain. Nevertheless, the analysis still demonstrated that the Major Clinical Response endpoint was associated with a greater expected incremental QALY gain than the Improvement endpoint, and future analyses can account for this parameter uncertainty appropriately by sampling values from a relevant statistical distribution. A third limitation was that most patients in the estimation sample were treated with rituximab at baseline, which may challenge the generalizability of the findings to patients who receive other treatments. The use of rituximab for SLE varies worldwide⁴³ and new treatments are likely to enter the healthcare system soon.⁴⁴ It is plausible that the incremental improvement in health state utility associated with achieving either composite endpoint is independent of the treatment prescribed at baseline. This assumption can be investigated within cohorts who do not receive biologic treatment or when

new treatments receive marketing authorization for SLE and their data become available in the national patient register.

Future research could replicate this analysis by using the EQ-5D-3L tariff for other countries to value the health gain of achieving the composite response endpoints in different decision-making jurisdictions. The methods for this study could also be used to estimate the QALY and monetary benefit of response for existing composite response endpoints in SLE,^{17,18} from the completed clinical trial data sets, to compare the relative health gains of achieving these different endpoints. Similarly, these methods can be applied when developing a bespoke composite response endpoint for other complex multisystem diseases to ensure that it reflects improvements in health state utility and outcomes that are valuable for decision makers. The cost of strategies to achieve these endpoints could also be compared with the monetary benefit of response in a full economic evaluation to estimate the net monetary benefit of treatment.³² Finally, after the introduction of value sets for the EQ-5D-5L,⁴⁵ it will be possible to estimate the incremental improvement in health gain associated with achieving all composite response endpoints for SLE and compare the findings with the estimated results based on the EQ-5D-3L value sets.

Conclusion

Composite response endpoints are becoming more common to determine whether patients with multisystem diseases respond to treatment within clinical trials and observational studies. HTA agencies and decision makers require empirical evidence to establish whether these endpoints correspond with improved health gain. Increases in health state utility, QALYs, and monetary benefit of response provide a way to estimate this health gain and demonstrate the usefulness of these endpoints for HTA. Establishing the value of health gains associated with composite response endpoints, in addition to their clinical significance, will ensure they correspond with outcomes most relevant to decision makers, patients, and improving the cost-effectiveness of care.

Article and Author Information

Accepted for Publication: July 11, 2022

Published Online: August 23, 2022

doi: <https://doi.org/10.1016/j.jval.2022.07.001>

Author Affiliations: Manchester Centre for Health Economics, Division of Population Health, Health Services Research and Primary Care, School of Health Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester Academic Health Science Centre, Manchester, England, UK (Gavan, Payne); Centre for Epidemiology Versus Arthritis, Centre for Musculoskeletal Research, Division of Musculoskeletal and Dermatological Sciences, School of Biological Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester Academic Health Science Centre, Manchester, England, UK (Bruce); NIHR Manchester Biomedical Research Centre, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, England, UK (Bruce).

Correspondence: Sean P. Gavan, PhD, Manchester Centre for Health Economics, 4.312 Jean McFarlane Building, Division of Population Health, Health Services Research and Primary Care, School of Health Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Oxford Rd, Manchester, England M13 9PL, United Kingdom. Email: sean.gavan@manchester.ac.uk

Author Contributions: *Concept and design:* Gavan, Bruce, Payne

Acquisition of data: Bruce

Analysis and interpretation of data: Gavan, Bruce, Payne

Drafting of the manuscript: Gavan

Critical revision of the paper for important intellectual content: Gavan, Bruce, Payne

Statistical analysis: Gavan, Payne

Provision of study materials or patients: Bruce

Obtaining funding: Bruce, Payne

Conflict of Interest Disclosures: Dr Gavan reported receiving grants from the Medical Research Council and Lupus UK during the conduct of the study and personal fees from UCB outside the submitted work. Dr Bruce reports receiving grants from the Medical Research Council and the National Institute for Health and Care Research during the conduct of the study, grants and speaker fees from GSK, consultancy and speaker fees from Astra Zeneca, and consultancy fees from UCB, BMS, Eli Lilly, and Aurinia outside the submitted work. Dr Payne reports receiving personal fees from UCB outside the submitted work. No other disclosures were reported.

Funding/Support: This research was funded by the Medical Research Council, United Kingdom as part of the MAXimizing Sle Therapeutic Potential by Application of Novel and Stratified approaches (MASTERPLANS) Consortium (MR/M01665X/1) and part funded by a grant from LUPUS UK. BILAG-BR has been funded by unrestricted educational donations from Roche, GSK, and LUPUS UK.

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Acknowledgment: The authors thank Prof Caroline Gordon, Prof Roger Levy, and Jane Dunnage for feedback on an earlier draft. The MASTERPLANS Consortium comprises The University of Manchester (Ian Bruce, Katherine Payne, Mark Lunt, Niels Peek, Nophar Geifman, Sean Gavan, Gillian Armit, Patrick Doherty, Jennifer Prattle, Narges Azadbakht, Angela Papazian, Helen Le Sueur, Carmen Farrelly, Clare Richardson, Zunaira Shabbir, and Lauren Hewitt), University of Bath (Neil McHugh), University of Birmingham (Caroline Gordon, John Reynolds, and Stephen Young), University of Cambridge (David Jayne, Vern Farewell, and Li Su), Imperial College London (Matthew Pickering, Elizabeth Lightstone, Alyssa Gilmore, and Marina Botto), King's College London (Timothy Vyse, David Lester Morris, and David D'Cruz), University of Leeds (Edward Vital, Miriam Wittmann, and Paul Emery), University of Liverpool (Michael Beresford, Christian Hedrich, Angela Midgley, and Jenna Gritzfeld), University College London (Michael Ehrenstein, David Isenberg, and Mariea Parvaz), and Patient and Public Involvement Group (Jane Dunnage, Jane Batchelor, Elaine Holland, and Pauline Upsall). The BILAG-BR study coordinator is Emily Sutton and the BILAG-BR project administrator is Alison Fountain. The BILAG-BR contributing centers and principal investigators comprise Aberdeen Royal Infirmary (Dr Hazem Youssef), Addenbrooke's Hospital (Dr David Jayne), Alder Hey Children's Hospital (Dr Liza McCann), Barnet and Chase Farm Hospital (Dr Rapti Mediwake), Basildon Hospital (Dr Anurag Bharadwaj), Chapel Allerton Hospital (Dr Ed Vital), City Hospital Birmingham - Sandwell and West Birmingham Hospitals NHS Trust (Professor Caroline Gordon), Countess of Chester Hospital (Dr Deepti Kapur), Doncaster Royal Infirmary (Prof Chee-Seng Yee), Freeman Hospital Newcastle (Dr Bridget Griffiths), George Eliot Hospital NHS Trust (Dr Abid Yusuf), Great Western Hospital (Dr Rosemary Waller), Guy's Hospital (Dr David D'Cruz), Hairmyres Hospital (Asad Zoma), Haywood Hospital (Dr Erin Vermaak), Imperial College Hospital (Dr Francesco Carlucci), Ipswich Hospital (Dr Richard Watts), King's College Hospital (Patrick Gordon), Leicester Royal Infirmary (Dr Shireen Shaffu), Lister Hospital (Dr Jananath Wijeyekoon), Liverpool University Hospitals NHS Foundation Trust (Dr Zoe McLaren), Llandudno Hospital (Dr Yasmeen Ahmad), Maidstone Hospital (Mike Batley), Manchester Royal Infirmary (Professor Ian Bruce), Musgrove Park Hospital (Dr Luke Gompels), New Cross Hospital and Cannock Chase Hospital (Dr T Sheeran), Norfolk and Norwich University Hospital (Dr Cee Yi Yong), Northampton General Hospital (Dr Rachel Jeffery), Northwich Park Hospital (Dr Shahir Hamdulay), Poole Hospital NHS Foundation Trust (Dr Fouz Rahmeh), Queen Alexandra Hospital Portsmouth (Dr Steven Young Min), Queen Elizabeth Hospital Birmingham (Ben Rhodes), Queen Elizabeth the Queen Mother Hospital, Margate (Dr Denise De Lord), Queens Medical Centre, Nottingham (Peter Lanyon), Royal Berkshire Hospital (Anoni Chan), Royal Blackburn Hospital (Professor Lee-Suan Teh), Royal Bournemouth and Christchurch Hospitals NHS Foundation Trust (Dr Jonathan Marks), Royal Cornwall Hospital (David Hutchinson), Royal Derby Hospital (Dr Marian Regan), Royal Devon and Exeter Hospital (Dr Richard Haigh),

Royal Free Hospital (Dr Richard Stratton), Royal Glamorgan Hospital (Dr Ceril Rhys-Dillon), Royal Hallamshire Hospital (Dr Mohamed Akil), Royal Liverpool Broadgreen Hospital (Dr Devesh Mewar), Royal United Hospitals Bath (Dr Sarah Skeoch), Russells Hall Hospital (Dr Nicola Erb), Salford Royal NHS Trust (Dr Edmond O'Riordan), Salisbury District Hospital (Dr Sarah Bartram), Solihull Hospital (Dr Mary Gayed), Southend Hospital NHS Trust (Bhaskar Dasgupta), Southmead Hospital Bristol (Dr Harsha Gunwardena), St Bartholomew's Hospital (Dev Pyne), St George's Hospital (Dr Arvind Kaul), St Helens and Knowsley Teaching Hospitals NHS Trust (Dr Madhu Mahindrakar), St Hellier Hospital (Dr Bhrigu Raj Sood), Tunbridge Wells Hospital (Mike Batley), University College Hospital (Professor David Isenberg), University Hospital Coventry & Warwickshire (Dr Nicola Gullick), University Hospital Southampton (Professor Christopher Edwards), University Hospitals Bristol NHS Foundation Trust (Dr Joanna C Robson), University Hospitals Plymouth NHS Trust (Dr Jon King), Worcestershire Acute Hospitals (Dr Adrian Farrell), Wythenshawe Hospital (Dr Sahena Haque), and Yeovil District Hospital (Dr Sally Knights).

REFERENCES

- Multiple endpoints in clinical trials: guidance for industry - draft guidance. Food and Drug Administration. <https://www.fda.gov/media/102657/download>. Accessed January 11, 2022.
- Guideline on multiplicity issues in clinical trials. European Medicines Agency. https://www.ema.europa.eu/en/documents/scientific-guideline/draft-guideline-multiplicity-issues-clinical-trials_en.pdf. Accessed January 11, 2022.
- Cordoba G, Schwartz L, Woloshin S, et al. Definition, reporting, and interpretation of composite outcomes in clinical trials: systematic review. *BMJ*. 2010;341:c3920.
- Guideline: endpoints used for relative effectiveness assessment - composite endpoints. EUnetHTA. https://www.eunetha.eu/wp-content/uploads/2018/03/composite_endpoints.pdf. Accessed January 11, 2022.
- Kennedy-Martin M, Slaap B, Herdman M, et al. Which multi-attribute utility instruments are recommended for use in cost-utility analysis? A review of national health technology assessment (HTA) guidelines. *Eur J Health Econ*. 2020;21(8):1245–1257.
- Brennan A, Akehurst R. Modelling in health economic evaluation. What is its place? What is its value? *Pharmacoeconomics*. 2000;17(5):445–459.
- Ferreira-González I, Permanyer-Miralda G, Busse J, et al. Methodologic discussions for using and interpreting composite endpoints are limited, but still identify major concerns. *J Clin Epidemiol*. 2007;60(7):651–662.
- Nestler-Parr S, Korchagina D, Toumi M, et al. Challenges in research and health technology assessment of rare disease technologies: report of the ISPOR Rare Disease Special Interest Group. *Value Health*. 2018;21(5):493–500.
- Ramaekers B, Joore M, Grutters J. How should we deal with patient heterogeneity in economic evaluation: a systematic review of national pharmacoeconomic guidelines. *Value Health*. 2013;16(5):855–862.
- Dall'Era M, Bruce I, Gordon C, Manzi S, McCaffrey J, Lipsky PE. Current challenges in the development of new treatments for lupus. *Ann Rheum Dis*. 2019;78(6):729–735.
- Arora S, Isenberg D, Castrejon I. Measures of adult systemic lupus erythematosus: disease activity and damage. *Arthritis Care Res (Hoboken)*. 2020;72(Suppl 10):27–46.
- Rauch G, Schüller S, Kieser M. *Planning and Analyzing Clinical Trials With Composite Endpoints*. Cham, Switzerland: Springer International Publishing AG; 2017.
- Claxton K, Sculpher M, McCabe C, et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. *Health Econ*. 2005;14(4):339–347.
- Touma Z, Gladman D. Current and future therapies for SLE: obstacles and recommendations for the development of novel treatments. *Lupus Sci Med*. 2017;4(1):e000239.
- Mullard A. FDA approves Astra Zeneca's anifrolumab for lupus. *Nat Rev Drug Discov*. 2021;20(9):658.
- Guidance for industry: systemic lupus erythematosus - developing medical products for treatment. Food and Drug Administration. <https://www.fda.gov/media/71150/download>. Accessed January 11, 2022.
- Wallace D, Kalunian K, Petri M, et al. Efficacy and safety of epratuzumab in patients with moderate/severe active systemic lupus erythematosus: results from EMBLEM, a phase IIb, randomised, double-blind, placebo-controlled, multicentre study. *Ann Rheum Dis*. 2014;73(1):183–190.
- Furie R, Petri M, Wallace D, et al. Novel evidence-based systemic lupus erythematosus responder index. *Arthritis Rheum*. 2009;61(9):1143–1151.
- Yee C, Farewell V, Isenberg D, et al. The BILAG-2004 index is sensitive to change for assessment of SLE disease activity. *Rheumatology (Oxford)*. 2009;48(6):691–695.
- Gladman D, Ibañez D, Urowitz M. Systemic lupus erythematosus disease activity index 2000. *J Rheumatol*. 2002;29(2):288–291.
- Petri M, Genovese M, Engle E, Hochberg M. Definition, incidence, and clinical description of flare in systemic lupus erythematosus. A prospective cohort study. *Arthritis Rheum*. 1991;34(8):937–944.
- Barber M, Drenkard C, Falasinnu T, et al. Global epidemiology of systemic lupus erythematosus [published correction appears in *Nat Rev Rheumatol*. 2021;17(10):642] *Nat Rev Rheumatol*. 2021;17(9):515–532.
- Fanouriakis A, Kostopoulou M, Alunno A, et al. 2019 update of the EULAR recommendations for the management of systemic lupus erythematosus. *Ann Rheum Dis*. 2019;78(6):736–745.
- MAXimizing Sle Therapeutic Potential by Application of Novel and Stratified approaches (MASTERPLANS). UK Research and Innovation. <https://gtr.ukri.org/projects?ref=MR%2FM01665X%2F1>. Accessed January 11, 2022.
- Clinical commissioning policy: rituximab for refractory systemic lupus erythematosus (SLE) in adults and post-pubescent children [200402p]. NHS England. <https://www.england.nhs.uk/wp-content/uploads/2020/07/Rituximab-for-refractory-Systemic-Lupus-Erythematosus-in-adults-and-post-pubescent-children-2.pdf>. Accessed January 11, 2022.
- Belimumab for treating active autoantibody-positive systemic lupus erythematosus: technology appraisal guidance [TA752]. National Institute for Health and Care Excellence. <https://www.nice.org.uk/guidance/ta752>. Accessed January 11, 2022.
- BILAG biologics register. The University of Manchester. <https://sites.manchester.ac.uk/bilag/>. Accessed January 11, 2022.
- EuroQol Group. EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy*. 1990;16(3):199–208.
- Dolan P. Modeling valuations for EuroQol health states. *Med Care*. 1997;35(11):1095–1108.
- Wing C, Simon K, Bello-Gomez R. Designing difference in difference studies: best practices for public health policy research. *Annu Rev Public Health*. 2018;39(1):453–469.
- Manca A, Hawkins N, Sculpher M. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. *Health Econ*. 2005;14(5):487–496.
- Stinnett A, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Med Decis Making*. 1998;18(2 Suppl):S68–S80.
- Guide to the methods of technology appraisal 2013: process and methods [PMG9]. National Institute for Health and Care Excellence. <https://www.nice.org.uk/process/pmg9/chapter/foreword>. Accessed January 11, 2022.
- StataCorp. Stata Statistical Software. Release 16.1. College Station, TX: StataCorp LLC; 2019.
- Rees F, Doherty M, Grainge M, Davenport G, Lanyon P, Zhang W. The incidence and prevalence of systemic lupus erythematosus in the UK, 1999–2012. *Ann Rheum Dis*. 2016;75(1):136–141.
- Population estimates for the UK, England and Wales, Scotland and Northern Ireland: mid-2020. Office for National Statistics. <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2020>. Accessed January 11, 2022.
- Wailoo A, Hernandez-Alava M, Manca A, et al. Mapping to estimate health-state utility from non-preference-based outcome measures: an ISPOR good practices for outcomes research task force report [published correction appears in *Value Health*. 2017;20(8):1226] *Value Health*. 2017;20(1):18–1227.
- Gavan S, Bruce I, Payne K. Generating evidence to inform health technology assessment of treatments for SLE: a systematic review of decision-analytic model-based economic evaluations. *Lupus Sci Med*. 2020;7(1):e000350.
- Furie R, Morand E, Bruce I, et al. What does it mean to be a British Isles Lupus Assessment Group-based Composite Lupus Assessment responder? Post hoc analysis of two phase III trials. *Arthritis Rheum*. 2021;73(11):2059–2068.
- Furie R, Petri M, Strand V, et al. Clinical, laboratory and health-related quality of life correlates of systemic lupus erythematosus responder index response: a post hoc analysis of the phase 3 Belimumab trials. *Lupus Sci Med*. 2014;1(1):e000031.
- Wolfe F, Michaud K, Li T, Katz RS. EQ-5D and SF-36 quality of life measures in systemic lupus erythematosus: comparisons with rheumatoid arthritis, noninflammatory rheumatic disorders, and fibromyalgia. *J Rheumatol*. 2010;37(2):296–304.
- Smith KGC, Jones RB, Burns SM, Jayne DRW. Long-term comparison of rituximab treatment for refractory systemic lupus erythematosus and vasculitis: remission, relapse, and re-treatment. *Arthritis Rheum*. 2006;54(9):2970–2982.
- Rydén-Aulin M, Boumpas D, Bultink I, et al. Off-label use of rituximab for systemic lupus erythematosus in Europe. *Lupus Sci Med*. 2016;3(1):e000163.
- Felten R, Dervovic E, Chasset F, et al. The 2018 pipeline of targeted therapies under clinical development for systemic lupus erythematosus: a systematic review of trials. *Autoimmun Rev*. 2018;17(8):781–790.
- Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727–1736.