

Évaluation thérapeutique dans le lupus érythémateux systémique

Raphaëlle Seror

Problèmes méthodologiques spécifiques à l'évaluation du LES et des maladies systémiques 209

Paramètres à mesurer 210

Mesure de l'activité du LES 211

Cas particuliers : scores spécifiques d'organes 217

Mesure des séquelles du LES 217

Utilisation en pratique 219

Conclusion 221

En préambule à ce chapitre, il est primordial d'expliquer les raisons qui rendent indispensables l'évaluation thérapeutique des maladies systémiques en général et du lupus érythémateux systémique (LES) en particulier. Ces dernières années, avec l'avènement de l'*evidence-based medicine* (ou médecine fondée sur des preuves), un effort considérable a été fourni par l'ensemble de la communauté scientifique afin d'étudier, d'améliorer et de standardiser les méthodes liées à l'évaluation thérapeutique. Dans le cas du LES, l'évaluation de l'effet des traitements a considérablement évolué, passant de simples études ouvertes monocentriques non contrôlées (séries de cas) à la réalisation d'essais multicentriques contrôlés randomisés. Ces efforts et ces progrès ont été rendus obligatoires par l'accroissement considérable de notre arsenal thérapeutique pour traiter ces maladies. Si l'arsenal était jusqu'au début des années 2000 limité aux corticoïdes et aux immunosuppresseurs « classiques », il a connu un essor considérable avec l'avènement des thérapeutiques immunologiques ciblées ou thérapies.

Ces avancées ont aussi été facilitées par le développement de critères de jugement valides, performants et adaptés. Effectivement, un des paramètres cruciaux ayant permis ces avancées dans le domaine de l'évaluation des maladies systémiques est la standardisation des études cliniques. Cette standardisation a été largement aidée par la création

de l'OMERACT. L'OMERACT, définie initialement comme *outcome measures in rheumatoid arthritis clinical trials*, puis élargie aux autres rhumatismes inflammatoires et maladies auto-immunes, est devenue *outcome measures in rheumatology* [1]. Il s'agit d'une initiative internationale créée en 1992 dont l'objectif était de travailler sur l'obtention de meilleurs outils d'évaluation dans le domaine des rhumatismes inflammatoires et des maladies auto-immunes. Elle comporte de nombreux groupes de travail incluant des méthodologistes et des experts de chacune de ces maladies, dont un travaillant spécifiquement sur le LES [2]. Ces avancées ont permis une meilleure définition des maladies (critères diagnostiques et de classification), de leur gravité (critères pronostiques), des malades à inclure dans les études (critères d'inclusion ou d'éligibilité) et des outils d'évaluation à utiliser (critères d'évaluation).

PROBLÈMES MÉTHODOLOGIQUES SPÉCIFIQUES À L'ÉVALUATION DU LES ET DES MALADIES SYSTÉMIQUES

Dans le contexte du LES et plus généralement des maladies systémiques, il est important de préciser un certain nombre d'éléments qui rendent leur évaluation beaucoup plus complexe que dans d'autres pathologies.

- La première spécificité est leur relative rareté, même si, au sein des maladies systémiques, le LES est une des plus fréquentes. Cette relative rareté pose le problème du recrutement des patients dans les études cliniques, que ce soit pour le développement et/ou la validation de critères d'évaluation ou pour l'évaluation de nouvelles thérapeutiques. Cette rareté rend d'autant plus cruciale l'obtention de critères de jugement valides et performants. Cette performance est mesurée à l'aide de différents paramètres appelés caractéristiques métrologiques. Les plus importantes sont la sensibilité au changement (leur capacité à détecter le changement au cours du temps) et la reproductibilité (leur fiabilité) [3, 4]. Ainsi, plus les qualités métrologiques d'un score sont bonnes, plus le nombre de sujets nécessaires à inclure dans les essais peut être diminué. L'obtention d'ou-

tils d'évaluation ayant de bonnes qualités métrologiques est effectivement indispensable pour permettre de ne pas rejeter à tort une thérapeutique potentiellement efficace.

- La deuxième spécificité est le caractère extrêmement polymorphe de cette maladie. En effet, par opposition aux rhumatismes inflammatoires qui sont des maladies le plus souvent « mono-organe », comme par exemple la polyarthrite rhumatoïde, le LES est par nature une maladie pouvant potentiellement toucher un grand nombre voire tous les organes. De ce fait, son évaluation est rendue beaucoup plus complexe et des scores d'activité simples tels que l'est le DAS 28 [5, 6] dans la polyarthrite rhumatoïde ne sont pas transposables au LES dont la présentation clinique est très variable.
- Enfin, l'absence de *gold standard* pour mesurer l'activité ou les séquelles est un autre problème qui se pose pour qui veut développer ou évaluer des critères de jugement. L'évaluation des qualités métrologiques se limite donc à la comparaison des scores développés et à l'avis d'experts qui, on le sait, est éminemment variable, selon l'expérience personnelle mais aussi l'importance que chaque expert a attribuée à chaque organe.

Ainsi, on conçoit bien que ces spécificités font du LES un champ d'investigation extrêmement complexe en termes de mesure.

PARAMÈTRES À MESURER

DÉFINITION DU CORE SET

Comme nous le verrons, pour cette maladie, il existe un grand nombre de paramètres à évaluer. Cependant, certains paramètres sont plus importants que d'autres à mesurer et la liste de ceux devant être impérativement évalués (et donc rapportés dans les essais cliniques portant sur le LES) a été définie par consensus au sein du groupe OMERACT spécifique. Cette « liste » minimale des paramètres d'intérêt s'appelle un *core set*. Dans le cadre du LES, le *core set* minimal inclut (tableau 21.1) : l'activité de la maladie, la sévérité (ou séquelles), la qualité de vie et les effets indésirables des traitements [2, 7, 8]. Nous reviendrons plus loin sur la définition de chacun de ces paramètres.

Pour cette maladie, de nombreux scores d'évaluation sont à notre disposition. Ces scores, même lorsqu'ils mesurent le même paramètre, ne sont pas tous identiques et diffèrent sur de nombreux points. Avant l'implémentation de ce *core set*, ils étaient utilisés de façon variable dans les études cliniques et certains paramètres essentiels étaient parfois oubliés, ce qui rendait difficile voire impossible la comparaison des études entre elles. L'implémentation du *core set* a été l'un des premiers pas vers la standardisation de l'évaluation du LES.

CONCEPTS D'ACTIVITÉ ET DE SÉVÉRITÉ

Le LES, bien qu'étant une maladie où le phénomène inflammatoire reste prépondérant, doit être appréhendé comme étant une maladie chronique, dont l'évolution est faite d'une alternance de poussées/rémissions. Les poussées de la maladie s'accompagnent généralement d'une majoration importante du processus inflammatoire pouvant toucher

Tableau 21.1. Domaines à prendre en considération pour inclusion dans les essais thérapeutiques ou les études observationnelles dans le LES (core set OMERACT IV).

Paramètres minimums devant être systématiquement inclus

- Activité de la maladie
- Séquelles
- Qualité de vie

Autres domaines à inclure conformément aux précédentes initiatives OMERACT

- Les coûts économiques y compris les coûts de santé
- Les événements indésirables

Autres domaines pouvant être pris en considération

- Le décès
- Le handicap
- La gravité de la maladie (en dehors de l'activité et des séquelles)
- La fatigue
- La fibromyalgie
- L'évaluation globale par le patient
- L'évaluation globale par le médecin
- L'hypertension
- Les mesures psychosociales
- Les données biologiques (auto-anticorps)
- Le statut de travail

un ou plusieurs organes et responsable de manifestations « aiguës ». Les poussées sont ensuite généralement suivies de périodes d'amélioration ou de rémission, obtenues parfois spontanément mais le plus souvent après la mise en route de traitements spécifiques non dénués d'effets secondaires (corticoïdes et/ou immunosuppresseurs). Sur certains organes, même une fois le processus inflammatoire (i. e. la poussée) maîtrisé, il peut persister des symptômes liés à des lésions tissulaires irréversibles ou séquelles, reflet de la sévérité de la maladie.

Ainsi, pour l'évaluation thérapeutique de cette maladie, dans la pratique clinique comme pour la conception de scores d'évaluation, il faut savoir distinguer et pouvoir mesurer deux concepts intriqués mais opposés que sont :

- l'activité de la maladie (ou *disease activity*) correspondant aux manifestations réversibles, liées à la poussée actuelle de la maladie. Dans le cadre de l'évaluation thérapeutique, les manifestations liées à l'activité sont les plus importantes à mesurer car ce sont celles sur lesquelles les traitements peuvent agir ;
- la sévérité ou les séquelles de la maladie (ou *damage*) doivent être distinguées des précédentes et correspondent aux lésions irréversibles liées aux poussées antérieures ou aux traitements. Ces manifestations ne peuvent, par définition, pas régresser. Il est donc inutile d'espérer les améliorer avec de nouveaux traitements immunosuppresseurs.

On voit qu'il est donc primordial de pouvoir mesurer séparément ces paramètres, en particulier afin de ne pas faire intervenir des manifestations irréversibles dans la mesure de l'activité [9]. En effet, l'utilisation d'un score ne faisant pas

le distingue entre activité et séquelles conduirait irrémédiablement à conclure, à tort ou à raison, à l'inefficacité des traitements. Cependant, la distinction entre activité et séquelles est bien souvent en pratique beaucoup plus difficile qu'il n'y paraît, les manifestations cliniques pouvant être identiques à ces 2 stades de la maladie. C'est le cas par exemple de certaines manifestations liées aux atteintes rénales (où le passage de l'insuffisance rénale aiguë réversible à la chronicité est parfois difficile à dater précisément) ou neurologiques (où après l'installation des symptômes, la récupération est souvent difficilement prévisible ; il est alors difficile de savoir à partir de quel moment on peut parler de séquelles irréversibles). Dans ces cas-là, c'est l'absence de régression des symptômes, malgré un traitement bien conduit et efficace, qui permet de faire cette distinction. Ainsi, lorsque ces manifestations deviennent chroniques et irréversibles, on peut alors parler de séquelles. La question encore incomplètement résolue est de savoir quel est le délai acceptable pour basculer de manifestations liées à l'activité à celles liées aux séquelles. En conséquence, la problématique de faire le distinguo entre activité de la maladie et séquelles est un paramètre primordial à prendre en compte dans la construction d'un score d'activité et lors de son utilisation en pratique clinique.

NOTION DE QUALITÉ DE VIE

Au cours de l'évolution de cette maladie, que ce soit au moment des poussées (du fait de la maladie ou des traitements) ou lors des périodes de rémission (du fait des séquelles), les patients peuvent voir leur qualité de vie altérée. Les paramètres conduisant à cette altération de la qualité de vie sont multiples et ne sont pas uniquement en rapport avec l'activité et la sévérité de la maladie [10, 11]. Ainsi, des paramètres comme le retentissement fonctionnel, psychologique, l'état de santé ou plus généralement la qualité de vie sont des paramètres qui doivent aussi être évalués dans le suivi de ces maladies, pour appréhender de façon complète leur impact [10].

Enfin, le LES peut être responsable d'un certain nombre de symptômes qui ne sont pas mesurables à l'aide de paramètres objectifs. Ces symptômes sont des manifestations cliniques ressenties par les patients, comme par exemple la douleur ou la fatigue. En métrologie, ces paramètres sont appelés critères subjectifs ou critères « mous ». Dans la littérature anglo-saxonne, ils sont appelés *patient-reported outcomes* (PROs) [11].

LIENS ENTRE CES PARAMÈTRES ET OBJECTIF DES TRAITEMENTS

Comme nous venons de le voir, l'évaluation du LES passe par la mesure de différents paramètres qui sont très différents les uns des autres. Certains d'entre eux, comme l'activité et la sévérité, font intervenir des éléments objectifs cliniques ou paracliniques et font appel aux connaissances et au raisonnement médical. Ces paramètres doivent donc être évalués par les cliniciens. Par contre, les critères « mous » ne sont mesurables que par le patient lui-même, car difficilement et imparfaitement appréhendés par les cliniciens [12].

Ainsi, il semble important d'être capable de mesurer indépendamment ces différents paramètres, tout en gardant à l'esprit qu'ils sont étroitement liés (figure 21.1). Ainsi, il apparaît

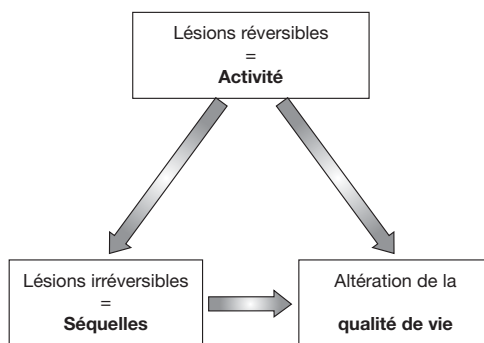


Figure 21.1. Liens entre activité, séquelles et qualité de vie.

clairement que la prise en charge des patients lupiques passe par le traitement des poussées pour diminuer l'activité de la maladie, pour en limiter ainsi les séquelles afin de minimiser les symptômes des patients et pour maintenir la qualité de vie. Cependant, on s'aperçoit en pratique que la prise en charge des poussées de la maladie est un des temps les plus importants car ce sont elles qui peuvent mettre en jeu le pronostic vital et être responsables des séquelles. Ainsi, au sein de l'ensemble des paramètres à mesurer, il apparaît que l'activité est tout de même le paramètre le plus important : c'est le *primum movens*, les autres paramètres en découlant.

MESURE DE L'ACTIVITÉ DU LES

Au sein des maladies systémiques, le LES est certainement la maladie pour laquelle la mesure de l'activité et les problèmes méthodologiques qui en découlent ont été les plus étudiés. C'est aussi la maladie systémique (polyarthrite rhumatoïde exclue) pour laquelle le plus grand nombre d'essais thérapeutiques ont été et sont toujours conduits. Ainsi, de nombreux scores d'activité ont été proposés : le BILAG (*british isles lupus assessment group index*) [13], le SLEDAI (*systemic lupus erythematosus disease activity index*) [14], le SLAM (*systemic lupus erythematosus activity measure*) [15], l'ECLAM (*european consensus lupus activity measurement*) [16, 17], le LAI (*lupus activity index*) [18] et le SLAQ (*systemic lupus activity questionnaire*) [19].

Parmi ces nombreux scores [15-18], on en distingue deux principaux : le BILAG [13] et le SLEDAI [14]. Bien que tous deux soient destinés à mesurer l'activité de la maladie, la « philosophie » et les méthodes ayant conduit à leur développement sont très différentes. Ces deux index ont subi depuis leur origine de nombreuses modifications. Nous détaillerons ici la « philosophie », les méthodes de développement et les principales modifications méthodologiques apportées à chacun de ces indices.

LE BILAG

Le BILAG a été le premier indice d'activité du lupus [13]. Développé en 1988 par l'équipe britannique de Birmingham, il est composé initialement de 86 items (étendus à 101 dans sa version la plus récente [20]) (tableau 21.2) couvrant

51. Pulmonary haemorrhage/vasculitis	
52. Interstitial alveolitis/pneumonitis	
53. Shrinking lung syndrome	
54. Aortitis	
55. Coronary vasculitis	
6. GASTROINTESTINAL	
56. Lupus peritonitis	
57. Abdominal serositis or ascites	
58. Lupus enteritis/colitis	
59. Malabsorption	
60. Protein losing enteropathy	
61. Intestinal pseudo-obstruction	
62. Lupus hepatitis	
63. Acute lupus cholecystitis	
64. Acute lupus pancreatitis	
7. OPHTHALMIC	
65. Orbital inflammation/myositis/proptosis	
66. Keratitis – severe	
67. Keratitis – mild	
68. Anterior uveitis	
69. Posterior uveitis/retinal vasculitis – severe	
70. Posterior uveitis/retinal vasculitis – mild	
71. Episcleritis	
72. Scleritis – severe	
73. Scleritis – mild	
74. Retinal/choroidal vaso-occlusive disease	
75. Isolated cotton-wool spots (cytoid bodies)	
76. Optic neuritis	
77. Anterior ischaemic optic neuropathy	
8. RENAL	
78. Systolic blood pressure (mmHg)	Value Y/N*
79. Diastolic blood pressure (mmHg)	Value Y/N*
80. Accelerated hypertension	Yes/No

81. Urine dipstick protein (+ = 1, ++ = 2, +++ = 3)	Y/N*
82. Urine albumin-creatinine ratio mg/mmol	Y/N*
83. Urine protein-creatinine ratio mg/mmol	Y/N*
84. 24 hours urine protein (g)	Value Y/N*
85. Nephrotic syndrome	Yes/No
86. Creatinine (plasma/serum) $\mu\text{mol/l}$	Y/N*
87. GFR (calculated) ml/min/1,73 m ²	Y/N*
88. Active urinary sediment	Yes/No
89. Active nephritis	Yes/No
9. HAEMATOLOGICAL	
90. Haemoglobin (g/dl)	Value Y/N*
91. Total white cell count ($\times 10^9/\text{l}$)	Value Y/N*
92. Neutrophils ($\times 10^9/\text{l}$)	Value Y/N*
93. Lymphocytes ($\times 10^9/\text{l}$)	Value Y/N*
94. Platelets ($\times 10^9/\text{l}$)	Value Y/N*
95. TTP	
96. Evidence of active haemolysis	Yes/No
97. Coombs' test positive (isolated)	Yes/No

9 domaines : signes généraux, atteinte cutanée, neurologique, articulaire, cardiaque, respiratoire, rénale et hémato­logique. Il évalue l'activité sur les 4 semaines précédant l'évaluation. Son développement (choix des items et définition des domaines) s'est basé sur un consensus d'experts. Chaque item est coté de manière transitionnelle en quatre classes selon que l'item est amélioré, stable, aggravé ou nouveau. C'est le seul indice dont la mesure de l'importance de l'activité de chaque domaine utilise comme *gold standard* l'intention de traiter. L'intention de traiter est ici différente et doit être distinguée du terme habituellement utilisé pour les essais thérapeutiques. Lors du développement du score, les experts ont défini, à partir de la cotation des items, pour chaque domaine, un niveau d'activité en « intention de traiter », c'est-à-dire correspondant au traitement que l'importance de l'atteinte du domaine en question nécessiterait selon eux. Le score de chaque domaine est ainsi obtenu par une combinaison complexe du score des items qui le composent. On obtient ainsi, pour chaque domaine, un score variant entre A (le plus haut niveau d'activité) et E (le plus faible niveau d'activité). Les définitions de ces niveaux d'activité sont données dans le tableau 21.3.

Les critères pour définir les poussées de la maladie ou la réponse au traitement sont définis dans le tableau 21.4. Ils s'appliquent à l'ensemble des domaines du score. Du fait de la complexité de ce score, ces critères sont essentiellement utilisés dans les essais thérapeutiques. Le BILAG est un excellent

Tableau 21.3. Niveaux d'activité des domaines du score BILAG : principe de « l'intention de traiter ».

Score BILAG	Définition
BILAG A	Maladie très active nécessitant des immunosuppresseurs ou une corticothérapie (> 20 mg/j)
BILAG B	Maladie active nécessitant une corticothérapie (< 20 mg/j) ou des anti-inflammatoires non stéroïdiens ou des antipaludéens de synthèse
BILAG C	Maladie stable nécessitant uniquement des traitements symptomatiques
BILAG D	Organe précédemment atteint actuellement inactif
BILAG E	Organe jamais atteint

score pour détecter les poussées [21-23]. La définition de la réponse au traitement, c'est-à-dire la disparition (sans apparition) de tous les A et B, est quant à elle un critère très strict, qui ne permet pas de détecter une amélioration si celle-ci n'est que partielle.

Depuis son développement initial, le BILAG n'est pas voué à être utilisé comme un score global. Ainsi, l'évaluation de l'activité s'effectue domaine par domaine (score de A à E) et un patient donné obtient un score de A, qu'il ait une ou plusieurs atteintes de niveau A. Certains auteurs ont proposé diverses conversions numériques [24, 25]. Celle proposée pour le BILAG 2004 est la suivante : A = 12, B = 8, C = 1 et D ou E = 0 [26]. Cependant, ces conversions numériques attribuent le même poids à chaque domaine, ne tenant pas compte de l'hétérogénéité existant entre les domaines et donc du caractère plus ou moins grave de certaines atteintes. De plus, aucune de ces conversions n'avait été validée, jusqu'à très récemment [25, 26].

LE SLEDAI

L'autre score dont l'utilisation est extrêmement répandue dans l'évaluation du lupus est le SLEDAI développé en 1992 [14]. Plusieurs modifications ont été suggérées, mais la plus utilisée est celle proposée pour l'essai du groupe SELENA sur l'utilisation du traitement hormonal substitutif

Tableau 21.4. Critères de réponse ou de poussée selon le score BILAG.

Critère de réponse	Définition
Poussée majeure	Nouveau score A
Poussée modérée	Passage d'un C, D ou E à un B
Réponse complète	Disparition (sans apparition) de tous les A et B
Réponse partielle	Disparition des A et ≤ 1 nouveau B

chez les patientes lupiques : le SELENA-SLEDAI [27] [28]. Les méthodes utilisées pour le développement de ce score sont totalement différentes de celles utilisées pour le BILAG. Il s'agit d'un score qui évalue globalement l'activité de la maladie. Il comporte 24 items, couvrant 9 domaines. Il mesure l'activité dans les 10 jours précédant l'évaluation. Comparativement au BILAG, le SLEDAI ne comptabilise pas de symptômes subjectifs tels que la fatigue qui constitue un des symptômes constitutionnels les plus fréquents au cours de la maladie, sans être nécessairement lié à une activité plus élevée de la maladie [29].

Lors de son développement, après une sélection initiale d'une longue liste d'items par des experts du LES, la sélection et la pondération des items ont été réalisées sur une large cohorte de patients. Cette étude de développement a utilisé l'évaluation globale de l'activité par le clinicien (PGA) comme *gold standard* (ou variable à expliquer) et les différents items comme variables explicatives. Ce processus a permis dans une analyse multivariée de sélectionner les items et de leur attribuer un poids. Contrairement au BILAG, les items sont cotés de façon binaire (c'est-à-dire présents ou absents). Toutefois, certains des items faisaient intervenir une notion d'évolutivité, l'item ne devant être pris en compte que s'il était nouveau ou aggravé. Le SELENA-SLEDAI est une version modifiée du SLEDAI (tableau 21.5). La description de certains paramètres est légèrement modifiée, mais les systèmes/organes et les scores pondérés sont les mêmes que pour le SLEDAI. Le SELENA-SLEDAI mesure la présence ou l'absence de signes cliniques, de symptômes ou d'anomalies de laboratoire du lupus. D'autres modifications du SLEDAI ont été proposées, le SLEDAI 2000 (ou SLEDAI-2K) en est une. Cette version a l'avantage de ne pas tenir compte de l'évolution des signes et des symptômes (amélioration ou aggravation) par rapport à la visite précédente [28]. Cette modification permet d'éviter un problème fréquent lié à ces items codés comme nouveaux ou aggravés. En effet, l'inclusion de tels items dans un score fait que ce dernier peut s'améliorer alors que l'état du patient est inchangé, du simple fait que le symptôme qui était présent à la première visite n'est plus pris en compte à une visite ultérieure puisqu'il n'est plus nouveau [10].

Enfin, contrairement au BILAG, la philosophie du SLEDAI est un peu différente : il s'agit d'une mesure de l'activité globale et le score ne peut donc pas être utilisé domaine par domaine. Le score global peut théoriquement varier de 0 à 105 (score d'un patient qui présenterait simultanément les 24 manifestations cliniques ou biologiques – situation inconcevable en pratique). De ce fait, un score ≥ 20 correspond déjà à une maladie lupique de très haute activité. Des seuils d'activité ont été proposés (tableau 21.6) [30], ainsi que les seuils définissant une amélioration ou une poussée de la maladie (tableau 21.7) [31-33].

Enfin, toujours dans le cadre de l'étude SELENA, des critères de poussée ont été définis (tableau 21.8), faisant intervenir différents paramètres, incluant le SELENA-SLEDAI, l'évaluation globale par le clinicien (PGA) et un certain nombre d'atteintes d'organes de gravité différente [27].

Bien que l'indice SELENA-SLEDAI soit utile pour juger de l'activité du LED, il est parfois difficile de connaître la gravité de la maladie chez un patient en fonction de son score total. En effet, deux patients ayant un score SELENA-SLEDAI de 8 peuvent avoir des manifestations extrêmement différentes.

Tableau 21.5. Score SLEDAI.

Score	Manifestations	Définition
8	Convulsion	Apparition récente. Exclusion des causes métaboliques, infectieuses ou médicamenteuses.
8	Psychose	Perturbation de l'activité normale en rapport avec une altération sévère de la perception de la réalité. Comprend : hallucinations, incohérence, appauvrissement du contenu de la pensée, raisonnement illogique, comportement bizarre, désorganisé ou catatonique. Exclusion d'une insuffisance rénale ou d'une cause médicamenteuse.
8	Atteinte cérébrale	Altération des fonctions mentales avec troubles de l'orientation, de la mémoire ou autre, d'apparition brutale et d'évolution fluctuante. Comprend : troubles de la conscience avec réduction des capacités de concentration, incapacité à rester attentif avec en plus 2 au moins des manifestations suivantes : troubles perceptifs, discours incohérent, insomnie ou somnolence diurne, augmentation ou diminution de l'activité psychomotrice.
8	Troubles visuels	Atteinte rétinienne du lupus. Comprend : nodules dysoriques, hémorragies rétiniennes, exsudats séreux ou hémorragies choroïdiennes, névrite optique. Exclusion d'une cause hypertensive, infectieuse ou médicamenteuse.
8	Nerfs crâniens	Neuropathie sensitive ou motrice d'apparition récente touchant un nerf crânien.
8	Céphalées	Céphalées sévères et persistantes, pouvant être migraineuses mais résistant aux antalgiques majeurs.
8	AVC	Accident vasculaire cérébral d'apparition récente. Artériosclérose exclue.
8	Vascularite	Ulcérations, gangrène, nodules digitaux douloureux, infarctus périunguéraux ou preuve histologique ou artériographie de vascularite.
4	Arthrites	Plus de 2 articulations douloureuses avec des signes inflammatoires locaux (douleur, tuméfaction ou épanchement articulaire).
4	Myosite	Douleur/faiblesse musculaire proximale associée à une élévation des CPK et/ou aldolases ou à des modifications électromyographiques ou à une biopsie montrant des signes de vascularite.
4	Cylindres urinaires	Cylindres de globules rouges.
4	Hématurie	> 5 g/champ en l'absence de lithiase, d'infection ou d'une autre cause.
4	Protéinurie	> 0,5 GR/24 heures. Apparition récente ou majoration récente de plus de 0,5 g/24 heures.
4	Pyurie	> 5 GB/champ en l'absence d'infection.
2	Nouveau rash	Apparition récente ou récurrence d'un rash cutané inflammatoire.
2	Alopécie	Apparition récente ou récurrence d'une alopécie en plaques ou diffuse.
2	Ulcères muqueux	Apparition récente ou récurrence d'ulcérations orales ou nasales.
2	Pleurésie	Douleur thoracique d'origine pleurale avec frottement ou épanchement ou épaississement pleural.
2	Péricardite	Douleur péricardique avec au moins l'une des manifestations suivantes : frottement, épanchement ou confirmation électrographique ou échographique.
2	Baisse du complément	Diminution du CH50, du C3 ou du C4 < la normale inférieure du laboratoire.
2	Élévation des anti-ADN	Positivité > 25 % par le test de Farr ou taux > la normale du laboratoire.
1	Fièvre	> 38 °C en l'absence de cause infectieuse.
1	Thrombopénie	< 100 000 plaquettes/mm ³ .
1	Leucopénie	< 3 000 GB/mm ³ en l'absence de cause médicamenteuse.

L'un pourrait avoir plusieurs manifestations légères (score 1 ou 2 pour chacune), l'autre une seule manifestation très grave telle qu'une atteinte neurologique (score 8). Cependant, chez un même individu, au cours du temps, il est généralement capable de détecter les modifications d'activité. Cette

dernière situation est celle des essais cliniques mais aussi celle à laquelle le clinicien est confronté au quotidien. Enfin, même si le score total peut rester identique alors même que la combinaison de symptômes a considérablement changé, cette situation reste peu fréquente.

Tableau 21.6. Niveau d'activité défini selon le score SLEDAI.

Score SLEDAI	Niveau d'activité
SLEDAI = 0	Pas d'activité
SLEDAI entre 1 et 5	Activité légère
SLEDAI entre 6 et 10	Activité moyenne
SLEDAI entre 11 et 19	Activité élevée
SLEDAI \geq 20	Très haute activité

Tableau 21.7. Critères de réponse ou de poussée selon le score SLEDAI.

Critère de réponse	Définition
Poussée sévère	Augmentation \geq 12 points du SLEDAI
Poussée modérée	Augmentation \geq 3 points du SLEDAI/SELENA-SLEDAI
Amélioration	Diminution \geq 4 points du SELENA-SLEDAI
Rémission	SLEDAI/SELENA-SLEDAI = 0

Tableau 21.8. Définition des poussées dans l'étude SELENA.

Poussée légère à modérée	Poussée sévère
1 ou plusieurs éléments suivants : <ul style="list-style-type: none"> ■ Une variation du score SLEDAI \geq 3 points <i>ou</i> ■ Nouvelle atteinte ou aggravation d'une atteinte cutanée, stomatite, sérite, arthrite, fièvre <i>ou</i> ■ Augmentation de la dose de prednisolone (< 0,5 mg/kg/j) <i>ou</i> ■ Ajout d'un AINS <i>ou</i> antimalarique <i>ou</i> ■ Augmentation \geq 1,0 sur l'échelle PGA 	1 ou plusieurs éléments suivants : <ul style="list-style-type: none"> ■ Une variation du score SLEDAI > 12 points <i>ou</i> ■ Nouvelle atteinte ou aggravation d'une atteinte du système nerveux central, vascularite, myosite, néphrite, plaquettes < 60 000, anémie hémolytique avec Hb < 7 mg/dl, nécessitant un doublement de la dose <i>ou</i> > 0,5 mg/kg/j de prednisolone, <i>ou</i> une hospitalisation pour le LES <i>ou</i> ■ Ajout d'un immunosuppresseur ■ Augmentation \geq 2,5 sur l'échelle PGA

L'ÉVALUATION GLOBALE DU CLINICIEN (PGA)

L'objectif de ces différents scores étant de se rapprocher de l'évaluation que pourrait faire un expert de la maladie, vu la complexité pour obtenir une mesure d'activité « parfaite », la

question souvent posée est de savoir pourquoi ne pas utiliser, justement, cette évaluation globale de l'activité par le clinicien à l'aide d'une échelle visuelle analogique. Cette question est bien évidemment pertinente, puisque cette échelle est simple et qu'elle est utilisée comme *gold standard* de l'évaluation de l'activité de la maladie à la fois dans les études de développement et/ou de validation de l'ensemble de ses scores. Cependant, il a été démontré dans plusieurs maladies systémiques que cette mesure n'était pas reproductible et était extrêmement variable d'un individu à l'autre [31, 34], ces études faisant pourtant participer des experts reconnus des maladies étudiées. L'utilisation de cette échelle dans des mains moins expertes est donc encore plus sujette à caution et ne permet donc pas de recommander l'utilisation seule de ce type d'outils de mesure.

Cette échelle peut toutefois être utile chez un individu donné suivi par un clinicien donné. Ainsi, ont été aussi déterminés des niveaux d'activité (tableau 21.9) et des critères de poussée (tableau 21.10) à l'aide de ce type d'échelle [32, 35]. Ces critères ont été utilisés en combinaison avec les scores précédemment décrits afin de définir le critère d'évaluation dans certains essais, comme l'essai SELENA.

Tableau 21.9. Niveau d'activité défini selon la PGA.

Score PGA (sur 3)	Niveau d'activité
PGA = 0	Pas d'activité
PGA entre 0 et 1	Activité légère
PGA entre 2 et 2,5	Activité modérée
PGA \geq 2,5	Activité sévère mettant en jeu le pronostic vital

Tableau 21.10. Critères de poussée selon le score PGA.

Critère de réponse	Définition
Poussée sévère	Augmentation \geq 2,5 points
Poussée modérée	Augmentation \geq 1 point
Augmentation significative d'activité	Augmentation \geq 0,3 point

LE CHOIX DU SCORE

Comme nous venons de le voir, les 2 principaux scores d'activité du lupus sont bien différents. Même si en pratique clinique il apparaît clairement évident que le BILAG est inutilisable, alors que le SLEDAI peut l'être, la situation dans les essais thérapeutiques n'est pas aussi évidente.

En fait, chacun de ces scores présente des avantages et des inconvénients qui sont résumés dans le tableau (tableau 21.11). Ainsi, il apparaît que le BILAG est le score le plus adapté à la détection des poussées, alors que le SLEDAI

Tableau 21.11. Forces et faiblesses du BILAG et du SLEDAI.

BILAG	SLEDAI
Forces	Faiblesses
<ul style="list-style-type: none"> ■ Évaluation organe par organe ■ Sensible au changement organe par organe (items nouveaux/aggravés) 	<ul style="list-style-type: none"> ■ Ne permet pas d'analyse précise organe par organe ■ Moins sensible au changement par organe (items présents/absents)
Faiblesses	Forces
<ul style="list-style-type: none"> ■ Pas de score global : difficile à utiliser comme seul critère de réponse car difficulté à détecter une amélioration partielle ■ Complexité + + + ■ Approche « intention de traiter » : a une validité limitée dans le temps avec les progrès thérapeutiques ■ Développé sur avis d'expert 	<ul style="list-style-type: none"> ■ Obtention score global : <ul style="list-style-type: none"> – facile d'utilisation dans les essais – détecte amélioration globale partielle ■ Plus facile à coter ■ Approche « évaluation globale » : invariable avec les progrès thérapeutiques ■ Développé sur données de malades

est plus performant pour détecter une amélioration, même incomplète, de l'activité de la maladie.

CAS PARTICULIERS : SCORES SPÉCIFIQUES D'ORGANES

Même si le plus souvent l'activité et le retentissement du LES se mesurent de façon globale avec les scores précédemment décrits, des scores spécifiques ont été développés, pour certains organes, afin de pouvoir mener des études cliniques ciblées, en particulier des essais thérapeutiques. C'est le cas par exemple du lupus cutané ou du lupus rénal.

LUPUS ÉRYTHÉMATEUX (LE) CUTANÉ

L'atteinte cutanée est probablement une des plus fréquentes au cours du LE. Son importance et son caractère polymorphe ont récemment été soulignés dans l'établissement des nouveaux critères de classification du LE [36]. En effet, elle comporte, de plus, de nombreux sous-types (lupus érythémateux cutané aigu, subaigu, chronique ou intermittent). De plus, ces différents sous-types ont un profil évolutif variable, certains pouvant entraîner des séquelles cicatricielles et d'autres non, rendant nécessaire à la fois l'évaluation de l'activité et l'évaluation des séquelles. Enfin, la mesure de l'importance d'une atteinte cutanée doit aussi faire intervenir d'autres paramètres tels que la surface corporelle mais aussi la localisation des atteintes. Effectivement, une atteinte étendue et/ou touchant des zones découvertes telles que le visage ou les mains aura certainement un impact plus important que des atteintes sur des zones couvertes et/ou de petite taille.

Ainsi, on comprend aisément que des scores globaux ne permettent pas une évaluation fine de ces atteintes cutanées. Ils ne sont donc pas adaptés à une utilisation dans ces essais spécifiquement dédiés aux patients atteints de lupus érythémateux cutané. C'est pour cette raison que des scores spécifiques ont été récemment développés : le *cutaneous lupus erythematosus disease area and severity index* (CLASI) [37] et sa version révisée (RCLASI) [30] (figure 21.2). Ce score permet une évaluation de tous les différents sous-types de lupus érythémateux cutané, y compris les atteintes muqueuses et l'alopécie, leur étendue et leur localisation. De plus, il évalue à la fois les atteintes liées à l'activité et celles liées aux séquelles, et ceci de façon séparée.

LUPUS ÉRYTHÉMATEUX RÉNAL

L'atteinte rénale est elle aussi très fréquente, touchant près de 50 % des patients lupiques. Cette atteinte a bien sûr un pronostic potentiellement péjoratif. De nombreux essais thérapeutiques dans le LED sont exclusivement dédiés aux patients atteints de néphropathie lupique. Même si le « meilleur » outil d'évaluation de l'activité de la néphropathie est l'histologie rénale, on comprend aisément que cet examen ne peut être répété à l'infini. Différents critères de réponses ont été utilisés dans les différents essais thérapeutiques rendant difficile les comparaisons. De plus, les scores d'activité globale ne sont pas adaptés à l'évaluation seule de l'atteinte rénale. Le BILAG et les autres scores de réponses utilisables par domaine pourraient être utilisés. Toutefois, ils semblent discordants et non parfaitement reproductibles dans l'évaluation de la réponse rénale [33]. C'est pour cette raison qu'un index de réponse rénale a été développé par le groupe SLICC (« Systemic lupus international collaborating clinics »), groupe de recherche international sur le lupus. Cet index a pour but de grader la réponse de façon homogène entre les études en 4 classes : réponse complète, partielle, activité identique ou aggravation. Sur des études préliminaires, il semble reproductible mais doit être encore évalué dans des études prospectives [7].

MESURE DES SÉQUELLES DU LES

Les objectifs d'un score de séquelles sont bien différents de ceux d'un score d'activité. Le score de séquelle a pour objectif de capturer toutes les atteintes survenues après le diagnostic du LES et responsables de manifestations irréversibles.

Pour le LES, un seul score de séquelles existe, il s'agit du SLICC/SDI (*SLE damage index*) [38]. Il comporte 41 items couvrant 12 domaines. Le score des items varie de 1 à 3 (tableau 21.12). Le score total varie de 0 à 46 points. Il a été développé à partir d'un consensus d'experts. Pour que les items soient comptabilisés, ils doivent être présents depuis plus de 6 mois ou ne pas être réversibles (comme par exemple un infarctus du myocarde, qui peut être comptabilisé immédiatement sans avoir à attendre le délai de 6 mois).

Contrairement aux scores d'activité, dans ce type de score, les atteintes sont comptabilisées, qu'elles soient ou non attribuables au SLE, comptent ainsi les effets secondaires du traitement. Par définition, le score est de 0 au diagnostic et seuls comptent les événements survenus après le diagnostic du LES. De plus, les items présents une fois dans l'histoire du LES sont comptabilisés à vie, mêmes si les symptômes qui y

Figure 21.2. RCLASI : Revised cutaneous lupus erythematosus disease area and severity index (d'après Kuhn *et al.* [30]).
Seuls le score cutané est illustré ici. Pour les scores muqueux et d'alopecie, le lecteur peut se référer à la publication princeps.

RCLASI : Sélectionnez le score de chaque lésion cutanée lupique la plus sévère dans chaque site

Date / / Score Date de naissance / / Sexe Homme Femme Nom du patient

LESION	ACTIVITE				LESIONS SEQUELLAIRES	
	Erythème	squame/ kératose	Œdème/ infiltration	Nodule hypodermique/ Plaque	Troubles de la pigmentation	Cicatrice/ atrophie
	0 Absent	0 Absent	0 Absent	0 Absent	0 Absent	0 Absent
	1 Rose pâle	1a Circonscrit annulaire/ lésion papulo- squameuse psoriasiforme	1 Modéré/ juste palpable	1 Induration sous-cutanée	1a Hypopigmentation	1 Cicatrice initiale
	2 Rouge	1b Squame adhérente circonscrite/ kératose folliculaire	2 Palpable et visible	2 Ulcération de nodule ou de plaque sous-cutanés	1b Hyperpigmentation	2a Sévère, ferme, atrophique, vermiculaire
	3 Rouge foncé, pourpre, violacé, croûteuse, hémorragique	2 Hyperkératose verruqueuse			2 Hypo et hyperpigmentation	2b lipoatrophie
LOCALISATION	Cuir chevelu					
	Oreilles					
	Nez, zones malaires					
	Lèvres					
	Le reste du visage					
	Décolleté					
	nuque et/ ou épaules					
	Thorax					
	Abdomen					
	Dos/fesses					
	Bras					
	Mains					
	Jambes					
	Pieds					
Somme de chaque lésion				Somme de chaque lésion		
Score d'activité				Score cicatriciel		

sont liés ont disparu (par exemple une ostéonécrose guérie ou bien une cataracte opérée). Enfin, certains items peuvent être comptabilisés 2 fois pour les événements répétés, tels que les accidents vasculaires ou l'ostéonécrose aseptique. Toutefois, ils ne sont comptabilisés plusieurs fois que s'ils surviennent à plus de 6 mois d'intervalle et non s'ils surviennent simultanément. Ainsi, pour compléter le score, cela nécessite de connaître l'intégralité de l'histoire du patient.

De plus, le score ne peut s'améliorer.

Cet outil mêle ainsi de nombreux concepts et mesure plus l'impact global de la maladie (et de son traitement) sur le patient atteint du LES. Il s'agit donc d'un outil mesurant l'impact d'événements passés de la maladie. Ne pouvant pas par définition s'améliorer, il ne peut donc être utilisé comme critère d'évaluation dans les études cliniques, mais il a démontré à plusieurs reprises qu'il avait une bonne valeur pronostique

Tableau 21.12. SLICC/SDI (SLE damage index) [38].**Item****Ocular (either eye, by clinical assessment)**

- Any cataract ever
- Retinal change or optic atrophy

Neuropsychiatric

- Cognitive impairment (e. g. memory deficit, difficulty with calculation, poor concentration, difficulty in spoken or written language, impaired performance levels) or major psychosis
- Seizures requiring therapy for 6 months
- Cerebrovascular accident ever (score 2 if > 1)
- Cranial or peripheral neuropathy (excluding optic)
- Transverse myelitis

Renal

- Estimated or measured glomerular filtration rate < 50 %
- Proteinuria $\geq 3,5$ gm/24 hours
- or
- End-stage renal disease (regardless of dialysis or transplantation)

Pulmonary

- Pulmonary hypertension (right ventricular prominence, or loud P2)
- Pulmonary fibrosis (physical and radiograph)
- Shrinking lung (radiograph)
- Pleural fibrosis (radiograph)
- Pulmonary infarction (radiograph)

Cardiovascular

- Angina or coronary artery bypass
- Myocardial infarction ever (score 2 if > 1)
- Cardiomyopathy (ventricular dysfunction)
- Valvular disease (diastolic murmur, or systolic murmur > 3/6)
- Pericarditis for 6 months, or pericardiectomy

Peripheral vascular

- Claudication for 6 months
- Minor tissue loss (pulp space)
- Significant tissue loss ever (e. g. loss of digit or limb) (score 2 if > 1 site)
- Venous thrombosis with swelling, ulceration, or venous stasis

Gastrointestinal

- Infarction or resection of bowel below duodenum spleen, liver, or gall bladder ever, for cause any (score 2 if > 1 site)

- Mesenteric insufficiency
- Chronic peritonitis

Musculoskeletal

- Muscle atrophy or weakness
- Deforming or erosive arthritis (including reducible deformities, excluding avascular necrosis)
- Osteoporosis with fracture or vertebral collapse (excluding avascular necrosis)
- Avascular necrosis (score 2 if > 1)
- Osteomyelitis

Skin

- Scarring chronic alopecia
- Extensive scarring or panniculum other than scalp and pulp space
- Skin ulceration (excluding thrombosis) for > 6 months
- Premature gonadal failure
- Diabetes (regardless of treatment)
- Malignancy (exclude dysplasia) (score 2 if > 1 site)
- Stricture or upper gastrointestinal tract surgery ever

[39, 40]. On comprend ainsi aisément qu'à un instant donné, il n'y ait que peu ou pas de corrélation entre les scores d'activité et les scores de *damage* [24, 41, 42]. Ainsi, ces derniers seraient plus une mesure du reflet de l'activité cumulée du LES.

UTILISATION EN PRATIQUE**ESSAIS CLINIQUES OU PRATIQUE CLINIQUE ?**

Avant de voir comment ces scores sont utilisés, on se doit de distinguer deux situations dans lesquelles ils peuvent l'être. La première est celle pour laquelle ils ont été conçus : la recherche clinique et en particulier les essais thérapeutiques. Dans ce cadre-là, ces scores sont le plus souvent utilisés comme critères de jugement afin d'évaluer l'effet du traitement à l'essai. Ces scores ont été développés pour cela, afin de modéliser, standardiser et quantifier une évaluation habituellement effectuée en routine par le clinicien en charge du patient. Dans le contexte des essais thérapeutiques, le choix du critère principal d'évaluation est primordial. En effet, il est important de savoir que c'est à partir de ce critère qu'est calculé le nombre de sujets nécessaires et que se définit la puissance de l'essai. La deuxième situation est celle de la vraie vie où le critère d'évaluation peut permettre l'évaluation par le clinicien de l'état du patient de façon reproductible, ce qui peut être une aide dans le suivi des patients mais aussi pour le dialogue entre cliniciens afin d'avoir un langage commun.

LE SCORE PARFAIT EXISTE-T-IL ?

Avant d'aller plus loin, il est primordial de délimiter le « champ d'action d'un score d'évaluation », c'est-à-dire de bien préciser ce que l'on peut attendre d'un score d'évaluation mais aussi et surtout ce que l'on ne peut pas en attendre et quelles en sont les limites. Ainsi, il faut avoir en tête qu'un score d'évaluation ne peut pas être un outil parfait, qu'il ne permet pas à l'échelon individuel, face à un malade donné, d'évaluer de façon parfaite l'état du patient aussi bien que le ferait un clinicien spécialiste de la maladie, qu'il n'est pas capable de détecter les changements minimes et qu'il ne remplace pas l'œil expert du clinicien dans la prise de décision thérapeutique.

Toutefois, il doit permettre d'évaluer de façon reproductible l'état du patient, même par un clinicien non spécialisé. Cette évaluation doit permettre dans les études cliniques de comparer les patients entre eux et entre les études, de comparer le patient avec lui-même (au cours de son suivi) et ainsi de détecter les changements de son état quand la variation est cliniquement pertinente.

LE FILTRE OMERACT

Pour remplir tous ces objectifs, les qualités nécessaires d'un score ont été définies [21]. Ainsi, pour être approuvé par l'OMERACT, le score doit répondre à 3 critères, appelés le « filtre OMERACT » :

1. La validité : le score mesure-t-il ce que l'on a l'intention de mesurer ? Le résultat est-il non biaisé et pertinent ? Il s'agit des notions de validité de face, de contenu et de construit.
2. La capacité discriminante : la mesure permet-elle de distinguer les différentes situations d'intérêt soit à un moment donné (pour la classification et le pronostic), soit au cours du temps (pour mesurer le changement) ? Il s'agit des notions de fiabilité et de sensibilité au changement.

Il faut bien séparer 2 situations distinctes. La première est celle où l'on souhaite pouvoir distinguer, au sein d'une population donnée, des individus entre eux. La deuxième situation est celle qui nous intéresse ici : l'objectif n'est plus de distinguer différents individus entre eux mais bien d'évaluer chez un même individu les changements de son état dans le temps. On tentera ici de répondre à la question : « Mon patient s'est-il amélioré ou non ? » Ce paramètre est essentiel quand on s'intéresse à la mesure de l'efficacité d'un traitement ou d'une intervention. Dans ce cas, la qualité de l'outil d'évaluation se mesure par un paramètre appelé la sensibilité au changement (ou capacité à détecter le changement au cours du temps) [3, 4, 43].

3. Faisabilité : la mesure peut-elle être appliquée facilement, en termes de contraintes de temps, d'argent et d'interprétabilité ? Il s'agit ici de l'utilisation pragmatique de la mesure, et ce critère est indispensable dans le succès de l'utilisation d'un score au quotidien.

Ce dernier paramètre est surtout primordial si l'on souhaite que le score puisse être utilisé en pratique clinique. Si l'on prend l'exemple du BILAG, il est bien évident que ce type de score ne peut pas être utilisé en pratique clinique et ne remplit pas le critère de faisabilité, en termes de temps et de possibilité de calcul. Toutefois, ce paramètre de faisabilité doit aussi être pris en compte lorsque les scores sont destinés à être utilisés dans les essais, car on conçoit aisément que des scores extrêmement complexes risquent aussi de poser des

problèmes dans les essais, en termes de risque d'erreur et de reproductibilité. Si l'on reprend l'exemple du BILAG, son utilisation dans les essais nécessite pour les investigateurs de réaliser des sessions d'entraînement à la cotation de ce dernier afin de pouvoir participer aux essais.

SCORE GLOBAL OU CRITÈRES DE RÉPONSE

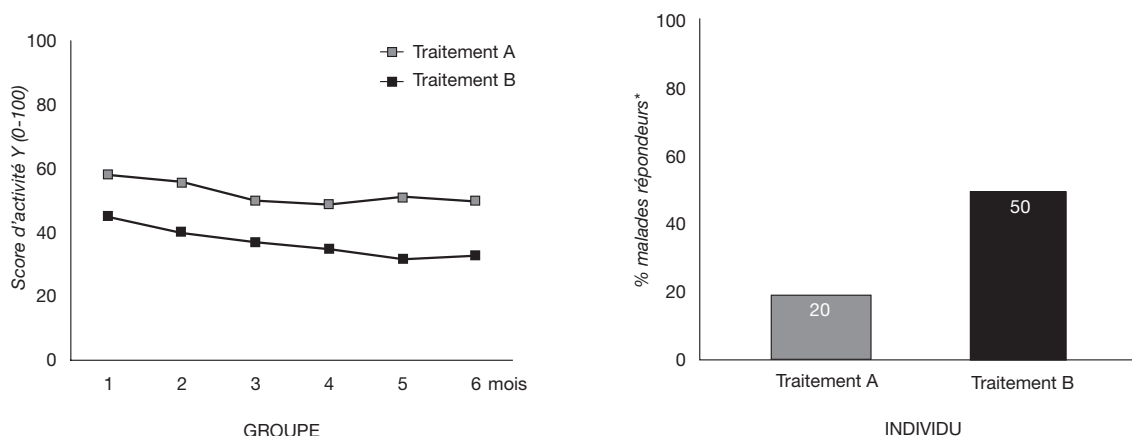
Une fois ces outils de mesure à disposition, l'obtention d'un score global au début et à la fin de l'étude ne résout pas totalement le problème puisqu'il faut encore déterminer comment les utiliser en pratique. Le mode de présentation des résultats est donc crucial. En effet, un résultat présenté sous forme de moyenne à l'échelon du groupe n'est pas facile d'interprétation pour un clinicien. Que penser d'un nouvel immunosuppresseur ayant permis de diminuer en moyenne 4 points du score SLEDAI chez 100 individus ? Cela nous est-il informatif pour en estimer le bénéfice pour un individu donné ? Ne préférons-nous pas savoir combien de ces 100 patients ont eu une diminution de 4 points ou plus (différence démontrée cliniquement pertinente) du SLEDAI, c'est-à-dire combien de ces 100 patients ont été améliorés ?

Sous-jacent à ces questions, c'est la distinction entre le « statistiquement significatif » et le « cliniquement pertinent » qui est mise en avant. C'est l'accent mis sur cette pertinence clinique qui a contribué à l'émergence de la notion de critères de réponse. Le principal intérêt de ces critères a été de transposer des résultats habituellement présentés à l'échelon du groupe en résultats présentés, et donc applicables, à l'échelon du patient. Ainsi, une information n'est plus présentée sous forme d'une diminution moyenne de score dans un groupe mais en terme de nombre de patients ayant diminué de X points du score Y, c'est-à-dire en nombre de patients répondeurs (figure 21.3).

Ce mode de présentation est déjà beaucoup plus facile d'interprétation mais se heurte à une difficulté supplémentaire qui est de définir le nombre de points X du score Y permettant de dire que le patient est répondeur. Ces définitions font intervenir le concept de différence minimale cliniquement pertinente (MCID : *minimal clinically important difference*, dans la littérature anglo-saxonne). Initialement développée pour les critères patients [44], elle s'est maintenant largement étendue aux critères d'activité [31, 45-49], afin de déterminer des seuils d'amélioration pour de nombreux scores d'activité et de définir la réponse au traitement. L'étape suivante a permis d'affiner cette mesure en faisant varier ces seuils pour définir non seulement les répondeurs, mais les bons et très bons répondeurs : c'est le principe des réponses EULAR [11, 49] ou ACR (20, 50 ou 70) [48] dans la polyarthrite rhumatoïde.

Enfin, comme nous l'avons vu, il existe pour le lupus plusieurs scores permettant de mesurer l'activité de la maladie. Ces scores ont des propriétés différentes, l'un pouvant être meilleur pour définir une poussée ou une rémission, c'est le cas du BILAG, l'autre pouvant être meilleur pour détecter l'amélioration, c'est le cas du SLEDAI. À la suite des essais ayant échoué à démontrer l'efficacité du rituximab dans le lupus érythémateux [50], traitement que la communauté scientifique trouvait efficace en pratique clinique [12, 51-53], des approches nouvelles ont émergé.

Ces approches visent à combiner les qualités de chacun de ces scores : l'aptitude du SLEDAI (*SLE disease activity index*) à détecter l'amélioration et celle du BILAG à détecter les poussées. Le score qui en découle est le *SLE responder index* (SRI) (cf. figure 21.4). Il a été mis au point au cours du développement



* Répondeur = Malade ayant une variation du score d'activité Y d'au moins X points (X = différence cliniquement pertinente)

Figure 21.3. Présentation des résultats : groupe versus individu.

clinique du belimumab (sur les données des essais de phase II [54, 55]), puis utilisé comme critère principal d'évaluation dans les essais de phase III (BLISS 52 et BLISS 76 [56]).

La réponse SRI (figure 21.4) se définit par :

- une diminution d'au moins 4 points du score SELENA-SLEDAI, qui sert à détecter l'amélioration ;
- l'absence de nouvelle atteinte BILAG A ou au maximum 1 nouvelle atteinte BILAG B, qui s'assure de l'absence d'aggravation ;
- l'absence d'aggravation significative de l'EVA médecin (moins de 0,3 point/3) par rapport à l'inclusion, qui permet de s'assurer que les deux précédents scores ne sont pas passés à côté d'une aggravation ou ne se sont pas faussement améliorés.

Ce score semble apte à détecter la réponse au traitement (figure 21.5) [54, 57]. De nouvelles données, encore non publiées, semblent confirmer la pertinence clinique potentielle de cet outil. En effet, à partir des données combinées de 2 essais de phase III (BLISS 52 et BLISS 76), les patients répondeurs ont été comparés aux non-répondeurs SRI, indépendamment du traitement reçu (placebo ou belimumab). Comparativement aux patients non répondeurs, les répondeurs avaient moins de poussées lupiques et pouvaient réduire de façon significativement plus importante leurs doses de corticoïdes. Ces données semblent donc montrer que les patients répondeurs SRI ont un bénéfice clinique global qui s'étend au-delà des paramètres inclus dans l'indice composite, renforçant sa pertinence clinique.

CONCLUSION

Pour conclure, on voit bien que l'évaluation des maladies systémiques comporte plusieurs paramètres étroitement liés entre eux tels que l'activité, les séquelles et la qualité de vie. Cependant, l'activité est le paramètre clé à mesurer, puisqu'il est à l'origine des 2 autres et qu'il est aussi le seul sur lequel on peut agir pour éviter l'apparition de séquelles et la dégradation de la qualité de vie.

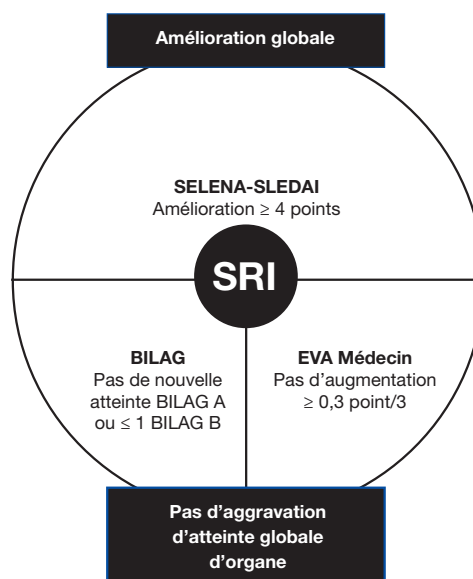


Figure 21.4. SLE responder index : nouvel index composite de réponse au traitement du lupus.

De plus, cette mesure est fondamentale pour concevoir des essais afin d'évaluer et de trouver de nouveaux traitements efficaces. Toutefois, la standardisation de cette mesure n'est pas chose aisée : comme nous l'avons vu, aucun score n'est parfait mais certains sont utiles. Les avancées récentes dans ce domaine utilisent des approches qui combinent plusieurs de ces scores. C'est le cas des nouveaux critères de réponse du lupus tels que le SRI, l'objectif étant ici d'exploiter les forces de chacun des scores afin de pallier leurs faiblesses, solution séduisante pour cette pathologie à la présentation clinique polymorphe et dont l'évaluation est donc complexe.