DEVELOPMENT AND VALIDATION OF THE BILAG-2004 INDEX FOR THE ASSESSMENT OF DISEASE ACTIVITY IN SYSTEMIC LUPUS ERYTHEMATOSUS

by

CHEE SENG YEE

A thesis submitted to
The University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

Rheumatology Research Group
Division of Infection and Immunology
School of Medicine
The University of Birmingham
February 2008

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

This research to develop and validate the BILAG-2004 index for the assessment of disease activity in patients with SLE had involved several validation studies.

The reliability of this index was shown in 2 exercises, as assessed by level of agreement, kappa statistic, ICC and disagreement between physicians. This study also highlighted the importance of training in ensuring optimal performance of the index. This index had construct validity as higher scores in the index were significantly associated increasing ESR, decreasing C3, decreasing C4, elevated anti-dsDNA antibody and increasing SLEDAI-2000 scores. Active disease scores were significantly associated with increase in therapy, confirming the criterion validity of the index. Sensitivity to change was demonstrated as changes in the score of the index were differentially related to change in therapy, with greater change in score having greater predictive power. Rasch analysis showed that this index had a good fit to the Rasch model, indicating that it is a unidimensional ordinal scale index with internal construct validity.

The validation studies led to revisions of the index to ensure it had face and content validity. The results of this research have shown that the revised BILAG-2004 index is valid for use to assess SLE disease activity.

Acknowledgements

I am deeply indebted to Caroline Gordon who has provided guidance, support and encouragement not only during this research project, but since the very beginning of my research career. She has been an excellent friend, colleague, supervisor and mentor. I am also very grateful to my other supervisors, David Isenberg and Vernon Farewell, for their advice, help and support.

It has been a pleasure working with members of British Isles Lupus Assessment Group (BILAG): David Isenberg, Caroline Gordon, Peter Maddison, Asad Zoma, Michael Snaith, Anisur Rahman, Munther Khamashta, David D'Cruz, Ian Bruce, Lee-Suan Teh, Mohammed Akil, Bridget Griffiths, Neil McHugh and Christopher Edwards. Their participation and contribution have been crucial in ensuring the success of this research. I am also grateful to Yasmeen Ahmad, Athiveeramapandian Prabu, Katharina Sokoll, Madelynn Chan and many other physicians from all the participating centres who have helped with the studies. Alan Tennant has provided invaluable advice on the Rasch analysis and Elizabeth Allen has helped with the calculation of the intraclass correlation coefficients in the reliability study. Dwomoa Adu and David Jayne have provided constructive advice on the renal aspect in the development of this index. Nurse specialists of the participating centres and Veronica Toescu have helped with the logistics and running of the various studies. I am also extremely grateful to Dagmar Scheel-Toellner for her advice and help with the preparation of this thesis.

This research has been a turning point in my career and it would not have been possible without Arthritis Research Campaign that provided the funding through the Clinical Research Fellowship scheme (Grant no 16081). I must also thank Lupus UK that has helped indirectly by providing funding support to nurse specialists and research assistant, and the

Wellcome Trust Clinical Research Facility in Birmingham for providing the infrastructure for the running of the research.

Table of Contents

ACKNOWLEDGEMENTS I

TABLE OF CONTENTS III

LIST OF FIGURES VII

LIST OF TABLES VIII

ABBREVIATIONS XIII

CHAPTER 1 INTRODUCTION 1

1.1 EPIDEMIOLOGY 1

1.2 IMMUNOPATHOLOGY 2
1.2.1 B-CELL ABNORMALITIES 3
1.2.2 T-CELL ABNORMALITIES 6
1.2.3 ABNORMALITIES IN APOPTOSIS 7
1.2.4 AUTOANTIBODIES AND IMMUNE COMPLEXES 8
1.2.5 COMPLEMENT SYSTEM ABNORMALITIES 9
1.2.6 CYTOKINE ABNORMALITIES 10

1.3 AETIOPATHOGENESIS 12
1.3.1 GENETIC FACTORS 12
1.3.2 ENVIRONMENTAL FACTORS 15
1.3.3 HORMONAL FACTORS 17

1.4 CLINICAL FEATURES 17

1.5 OUTCOMES IN SLE 21

1.6 TREATMENT 23

1.7 ASSESSMENT OF DISEASE ACTIVITY 25
1.7.1 BIOMARKERS 26
1.7.2 COMPOSITE CLINICAL DISEASE ACTIVITY INDICES 30
1.7.3 PROBLEMS WITH CURRENT INDICES 35

1.8 DEVELOPMENT AND VALIDATION OF THE BILAG-2004 INDEX 38
1.8.1 FACE AND CONTENT VALIDITY 42
1.8.2 RELIABILITY 43
1.8.3 CONSTRUCT AND CRITERION VALIDITY 46
1.8.4 SENSITIVITY TO CHANGE 47
1.8.5 PREDICTIVE VALIDITY 49
1.8.6 RASCH ANALYSIS 49

1.9 OBJECTIVES OF RESEARCH 52

CHAPTER 2 GENERAL METHODOLOGY 53

2.1 DEFINITION OF CHANGE IN THERAPY **54**

2.2 COMMON STATISTICAL ANALYSIS **57**

CHAPTER 3 RELIABILITY STUDY IN ROUTINE CLINICAL PRACTICE **59**

3.1 ABSTRACT **59**

3.2 OBJECTIVES **61**

3.3 INTRODUCTION **61**

3.4 METHODS **62**

3.5 ANALYSIS **63**

3.6 RESULTS **66**
3.6.1 INTER-RATER AGREEMENT AND RELIABILITY OF THE BILAG-2004 INDEX 67
3.6.2 INTER-RATER AGREEMENT OF SLEDAI-2000 INDEX 76
3.6.3 ABILITY TO DETECT DISEASE ACTIVITY 78

3.7 DISCUSSION **81**

CHAPTER 4 CROSS SECTIONAL STUDY **89**

4.1 ABSTRACT **89**

4.2 OBJECTIVES **91**

4.3 INTRODUCTION **91**

4.4 METHODS **92**

4.5 ANALYSIS **92**

4.6 RESULTS **94**
4.6.1 CONSTRUCT VALIDITY 96
4.6.2 CRITERION VALIDITY 100
4.6.3 OPTIMAL SLEDAI-2000 CUT-OFF SCORE FOR ACTIVE DISEASE 102
4.6.4 COMPARISON OF INDICES 103

4.7 DISCUSSION **104**

CHAPTER 5 FACE AND CONTENT VALIDITY OF RENAL SYSTEM **109**

5.1 ABSTRACT **109**

5.2 OBJECTIVE **110**

5.3 INTRODUCTION **110**

5.4 METHODS **111**

5.5 ANALYSIS **111**

5.6 RESULTS **113**

5.7 DISCUSSION **116**

CHAPTER 6 SENSITIVITY TO CHANGE STUDY 121

6.1 ABSTRACT 121

6.2 OBJECTIVES 123

6.3 INTRODUCTION 123

6.4 METHODS 124

6.5 ANALYSIS 124

6.6 RESULTS 126
6.6.1 SENSITIVITY TO CHANGE OF THE BILAG-2004 INDEX 127
6.6.2 SENSITIVITY TO CHANGE OF CLASSIC BILAG INDEX 129
6.6.3 SENSITIVITY TO CHANGE OF SLEDAI-2000 131
6.6.4 COMPARISON OF INDICES 138
6.6.5 RELATIONSHIP BETWEEN SLEDAI-2000 WITH THE BILAG-2004 AND CLASSIC BILAG INDICES 145

6.7 DISCUSSION 154

CHAPTER 7 RASCH ANALYSIS 162

7.1 ABSTRACT 162

7.2 OBJECTIVE 164

7.3 INTRODUCTION 164

7.4 METHODS 165

7.5 ANALYSIS 165

7.6 RESULTS 167

7.7 DISCUSSION 172

CHAPTER 8 FINAL DISCUSSION 175

8.1 FUTURE RESEARCH 183

8.2 CONCLUSION 186

APPENDIX 1 BILAG-2004 INDEX 187

APPENDIX 2 CLASSIC BILAG INDEX 211

APPENDIX 3 SLEDAI-2000 INDEX 226

APPENDIX 4 SLICC/ACR DAMAGE INDEX 228

APPENDIX 5 AGREEMENT TABLES FROM RELIABILITY STUDY 230

APPENDIX 6 REVISED BILAG-2004 INDEX (POST-RELIABILITY STUDY) 239

APPENDIX 7 REVISED BILAG-2004 INDEX (POST-LONGITUDINAL STUDY) 267

PUBLICATIONS 297

REFERENCES 298

List of Figures

Figure 1-1: Development of high affinity B-cells in a germinal centre (reproduced from Salmon and Gordon 1999). B, B-cell; CB, centroblast; CC, centrocyte; FDC, follicular dendritic cells; mB, memory B-cell; PC, plasma cell; T, T-cell.....4

Figure 1-2: The process of selecting memory B-cells in a germinal centre (reproduced from Salmon and Gordon 1999).....4

Figure 1-3: Typical receiver operating characteristic plot. 47

Figure 1-4: Representation of Rasch measurement using item-person map. (square represents person, circle represent item, shaded areas represent data that do not fit the Rasch model).....51

Figure 7-1: Threshold Map of Items in the BILAG-2004 Index after rescoring.....169

Figure 7-2: Person-Item threshold distribution of the BILAG-2004 index with Rasch analysis.....171

List of Tables

Table 1-1: Cumulative incidence of SLE manifestations (Adapted from Dubois' Lupus Erythematosus, 7th edition).....18

Table 1-2: The 1997 ACR revised criteria for classification of SLE. 20

Table 1-3: Interpretation of kappa statistic. 44

Table 1-4: Typical table of agreement. 45

Table 2-1: Dosing levels of medications used for defining change in therapy. 55

Table 3-1: Weighting used to calculate weighted agreement and weighted kappa statistic. 64

Table 3-2: Demographics of patients recruited into the reliability exercises.....66

Table 3-3: Distribution of disease activity across systems for Exercise 1 using the BILAG-2004 index as assessed by local physician.....66

Table 3-4: Distribution of disease activity across systems for Exercise 2 using the BILAG-2004 index as assessed by local physician.....67

Table 3-5: Agreement for each system in the BILAG-2004 index for Exercise 1.....68

Table 3-6: Kappa statistic (unweighted and weighted) and ICC of each system in the BILAG-2004 index for Exercise 1. 69

Table 3-7: Analysis of disagreement in scoring of the BILAG-2004 index between raters in Exercise 1. 70

Table 3-8: Agreement for each system in the BILAG-2004 index for Exercise 2.....72

Table 3-9: Kappa statistic (unweighted and weighted) and ICC of each system in the BILAG-2004 index for Exercise 2. 72

Table 3-10: Analysis of disagreement in scoring of the BILAG-2004 index between raters in Exercise 2. 73

Table 3-11: Analysis of level of agreement, kappa statistic, ICC and disagreements between raters in Exercise 1 for the Constitutional, Neuropsychiatric and Musculoskeletal systems before and after removal of some items (fatigue, lupus migraine/cluster/tension headache, mood disorder, anxiety disorder and arthralgia/myalgia).....75

Table 3-12: Analysis of level of agreement, kappa statistic, ICC and disagreements between raters in Exercise 2 for the Constitutional, Neuropsychiatric and Musculoskeletal systems before and after removal of some items (fatigue, lupus migraine/cluster/tension headache, mood disorder, anxiety disorder and arthralgia/myalgia).....76

Table 3-13: Level of agreement between raters for items in SLEDAI-2000 index.....77

Table 3-14: Cross tabulation of BILAG-2004 and SLEDAI-2000 scores for patients recruited into the study. 78

Table 3-15: Treatment analysis of patients with high BILAG-2004 and low SLEDAI-2000 scores with different cut-off scores used to define active disease with SLEDAI-2000. 79

Table 3-16: Treatment analysis of patients with low BILAG-2004 and high SLEDAI-2000 scores with different cut-off scores used to define active disease with SLEDAI-2000. 80

Table 4-1: Demographics of patients recruited into the cross sectional study (n=369). .. 95

Table 4-2: Distribution of disease activity based according to the BILAG-2004, Classic BILAG and SLEDAI-2000 indices (n=1510). 96

Table 4-3: Cross tabulation of overall BILAG-2004 scores with constructs (ESR, anti-dsDNA antibody, C3, C4 and SLEDAI-2000 score). 97

Table 4-4: Association of ESR, anti-dsDNA antibody, C3, C4 and SLEDAI-2000 scores respectively with higher overall BILAG-2004 scores. 99

Table 4-5: Association of ESR, anti-dsDNA antibody, C3, C4 and SLEDAI-2000 scores respectively with higher overall BILAG-2004 scores (when only first visit from each patient was used in the analysis). 100

Table 4-6: Cross tabulation of overall BILAG-2004 scores with change in therapy (n=1510). 101

Table 4-7: Association of increase in therapy with overall BILAG-2004 scores. 101

Table 4-8: Association of increase in therapy with overall BILAG-2004 scores (when only the first visit from each patient was used in the analysis). 101

Table 4-9: Receiver operating characteristic curves analysis of SLEDAI-2000 index... 102

Table 4-10: Receiver operating characteristic curves analysis of SLEDAI-2000 index using modified total score excluding the items anti-dsDNA antibody and complements. 103

Table 4-11: Sensitivity, specificity, positive predictive value and negative predictive value for increase in therapy of the BILAG-2004, Classic BILAG and SLEDAI-2000 indices. 104

Table 5-1: Cross tabulation of BILAG-2004 Renal system score (excluding Categories D and E) with change in therapy following assessment. 113

Table 5-2: Cross tabulation of modified BILAG-2004 Renal system score (in which urine protein dipstick result was superseded by other methods of urine protein estimation) with change in therapy following assessment. 114

Table 5-3: Cross tabulation of modified BILAG-2004 Renal system score (in which the threshold for definition of improvement in proteinuria was changed to 33% and urine protein dipstick result was superseded by other methods of urine protein estimation) with change in therapy following assessment. 115

Table 5-4: Cross tabulation of modified BILAG-2004 Renal system score (in which the threshold for definition of improvement in proteinuria was changed to 25% and urine protein dipstick result was superseded by other methods of urine protein estimation) with change in therapy following assessment. 115

Table 5-5: Cross tabulation of modified BILAG-2004 Renal system score (in which the threshold for definition of improvement in proteinuria was changed to 25%, additional Grade B criterion of urine protein excretion of at least 0.5g/day (or equivalent) that has not improved/decreased by 25%, and urine protein dipstick result would be superseded by other methods of urine protein estimation) with change in therapy following assessment. 116

Table 5-6: New criteria for proteinuria for Grades A, B and C of Renal system for the BILAG-2004 index (changes from the old criteria in italic) 119

Table 6-1: Demographics of patients recruited into the study (n=353). 126

Table 6-2: Cross tabulation of changes in disease activity (according to the BILAG-2004, Classic BILAG and SLEDAI-2000 indices) against change in therapy (n=1414). 127

Table 6-3: Sensitivity to change analysis of the BILAG-2004 index determining the association of change in overall BILAG-2004 score with change in therapy (n=1414).. 128

Table 6-4: Sensitivity to change analysis of the BILAG-2004 index after excluding change of Grade D/E to C from the definition of increase in activity (n=1414). 128

Table 6-5: Analysis on the subgroups of changes in overall BILAG-2004 score and its association with change in therapy (n=1414). 129

Table 6-6: Sensitivity to change analysis of Classic BILAG index determining the association of change in overall Classic BILAG score with change in therapy (n=1414). 130

Table 6-7: Sensitivity to change analysis of Classic BILAG index after excluding change of Grade D/E to C from the definition of increase in activity (n=1414). 130

Table 6-8: Analysis on the subgroups of changes in overall Classic BILAG score and its association with change in therapy (n=1414). 131

Table 6-9: Association between change in therapy with change in SLEDAI-2000 score and SLEDAI-2000 score of previous visit. 132

Table 6-10: Receiver operating characteristic curves analysis of SLEDAI-2000 index to determine the optimal increase in score associated with increase in therapy..... 133

Table 6-11: Receiver operating characteristic curves analysis of SLEDAI-2000 index to determine the optimal decrease in score associated with no increase in therapy..... 134

Table 6-12: Sensitivity to change analysis of SLEDAI-2000 index with Model A of change in score (minimal increase of 2 and minimal decrease of 2) (n=1414). 135

Table 6-13: Sensitivity to change analysis of SLEDAI-2000 index with Model B of change in score (minimal increase of 3 and minimal decrease of 1) (n=1414). 135

Table 6-14: Analysis on the subgroups of changes in SLEDAI-2000 score based on magnitude of change (Model C) and its association with change in therapy (n=1414).. 136

Table 6-15: Analysis on the subgroups of categorical changes in SLEDAI-2000 score (Model D) and its association with change in therapy (n=1414). 137

Table 6-16: Area under the receiver operating characteristic curve for the various models of change in SLEDAI-2000 score in explaining increase in therapy..... 137

Table 6-17: Comparison of the ability of change in BILAG-2004 score and change in Classic BILAG score in predicting change in therapy (n=1414).....139

Table 6-18: Sensitivity, specificity, positive predictive value and negative predictive value for the model of increase in activity, according to the BILAG-2004 and Classic BILAG indices, in predicting increase in therapy.....139

Table 6-19: Comparison of the ability of change in BILAG-2004 score and change in SLEDAI-2000 score (continuous variable) in predicting change in therapy (n=1414)..139

Table 6-20: Comparison of the ability of change in BILAG-2004 score and change in SLEDAI-2000 score (continuous variable) with previous visit SLEDAI-2000 score, in predicting change in therapy (n=1414).....140

Table 6-21: Comparison of the ability of change in BILAG-2004 score and Model A of change in SLEDAI-2000 score in predicting change in therapy (n=1414).....140

Table 6-22: Comparison of the ability of change in BILAG-2004 score and Model B of change in SLEDAI-2000 score in predicting change in therapy (n=1414).....141

Table 6-23: Comparison of the ability of change in BILAG-2004 score and Model C of change in SLEDAI-2000 score in predicting change in therapy (n=1414).....141

Table 6-24: Comparison of the ability of change in BILAG-2004 score and Model D of change in SLEDAI-2000 score in predicting change in therapy (n=1414).....142

Table 6-25: Comparison of the ability of change in Classic BILAG score and change in SLEDAI-2000 score (continuous variable) in predicting change in therapy (n=1414)..142

Table 6-26: Comparison of the ability of change in Classic BILAG score and change in SLEDAI-2000 score (continuous variable) with previous visit SLEDAI-2000 score, in predicting changes in therapy (n=1414).....143

Table 6-27: Comparison of the ability of change in Classic BILAG score and Model A of change in SLEDAI-2000 score in predicting changes in therapy (n=1414).....143

Table 6-28: Comparison of the ability of change in Classic BILAG score and Model B of change in SLEDAI-2000 score in predicting change in therapy (n=1414).....144

Table 6-29: Comparison of the ability of change in Classic BILAG score and Model C of change in SLEDAI-2000 score in predicting change in therapy (n=1414).....144

Table 6-30: Comparison of the ability of change in Classic BILAG score and Model D of change in SLEDAI-2000 score in predicting change in therapy (n=1414).....145

Table 6-31: Relationship between change in SLEDAI-2000 score (minimal change of 1) and change in BILAG-2004 score in explaining change in therapy (n=1414).....147

Table 6-32: Relationship between change in SLEDAI-2000 score (minimal change of 2) and change in BILAG-2004 score in explaining change in therapy (n=1414).....148

Table 6-33: Relationship between change in SLEDAI-2000 score (minimal change of 3) and change in BILAG-2004 score in explaining change in therapy (n=1414).....149

Table 6-34: Relationship between change in SLEDAI-2000 score (minimal change of 4) and change in BILAG-2004 score in explaining change in therapy (n=1414).....150

Table 6-35: Relationship between change in SLEDAI-2000 score (minimal change of 1) and change in Classic BILAG score in explaining change in therapy (n=1414).....151

Table 6-36: Relationship between change in SLEDAI-2000 score (minimal change of 2) and change in Classic BILAG score in explaining change in therapy (n=1414).....152

Table 6-37: Relationship between change in SLEDAI-2000 score (minimal change of 3) and change in Classic BILAG score in explaining change in therapy (n=1414).....153

Table 6-38: Relationship between change in SLEDAI-2000 score (minimal change of 4) and change in Classic BILAG score in explaining change in therapy (n=1414).....154

Table 7-1: Demographics of patients recruited into the study (n=353).....167

Table 7-2: Distribution of disease activity across 9 systems according to the BILAG-2004 index (n=353).....168

Table 7-3: Individual Item Fit of the BILAG-2004 Index with initial Rasch analysis...169

Table 7-4:Final Individual Item Fit of the BILAG-2004 Index. 170

Table 7-5: Distribution of disease activity across 9 systems according to the BILAG-2004 index in the validation sample (n=347).....171

Table 7-6: Individual Item Fit of the BILAG-2004 Index with the validation sample. .172

Table 8-1: Definition for inflammatory arthritis according to the BILAG-2004, Classic BILAG and SLEDAI-2000 indices.....179

Abbreviations

ACR	American College of Rheumatology
Anti-dsDNA	Anti-double-stranded DNA
AUC	Area under the curve
BAFF	B-cell activating factor
BILAG	British Isles Lupus Assessment Group
BILAG-2004	British Isles Lupus Assessment Group 2004 index
CI	Confidence intervals
Classic BILAG	Classic British Isles Lupus Assessment Group index
CRP	C-reactive protein
CTLA-4	Cytotoxic T lymphocyte antigen-4
DIF	Differential item functioning
DNA	Deoxyribonucleic acid
EBV	Epstein-Barr virus
ECG	Electrocardiograph
ECLAM	European consensus lupus activity measurement
ESR	Erythrocyte sedimentation rate
FcγR	Fc-gamma receptor
FcγRI	Fc-gamma receptor-I
FcγRII	Fc-gamma receptor-II
FcγRIIA	Fc-gamma receptor-IIA
FcγRIII	Fc-gamma receptor-III
FcγRIIIA	Fc-gamma receptor-IIIA
GFR	Glomerular filtration rate

HLA	Human leucocyte antigen
HRQOL	Health-related quality of life
ICC	Intra-class correlation coefficient
IgG	Immunoglobulin G
IgM	Immunoglobulin M
LAI	Lupus activity index
LAI-P	Lupus activity index in pregnancy
LFA-1	Lymphocyte function-associated antigen 1
MBL	Mannose-binding lectin
MHC	Major histocompatibility complex
NPV	Negative predictive value
NSAIDs	Non-steroidal anti-inflammatory drugs
OR	Odds ratio
PCA	Principal component analysis
PD-1	Programmed cell death-1
PGA	Physician's global assessment
PPV	Positive predictive value
ROC	Receiver operating characteristic
RNP	Ribonucleoprotein
SD	Standard deviation
SLAM	Systemic lupus activity measure
SLE	Systemic lupus erythematosus
SLEDAI	SLE disease activity index
SLEDAI-2000	SLE disease activity index 2000

SLICC	System Lupus International Collaborating Clinics
SLICC/ACR DI	Systemic Lupus International Collaborating Clinics/American College of Rheumatology damage index
TCR	T cell receptor
TNF	Tumour necrosis factor
TNF-α	Tumour necrosis factor α
WHO	World Health Organisation

Chapter 1 Introduction

Systemic lupus erythematosus (SLE) is a multi-system auto-immune disorder which is characterised by the production of numerous auto-antibodies. The term 'lupus' is a Latin word that literally means wolf. How this disease became associated with this animal is not known, but it has been suggested that the skin rash resembled a wolf's bite. For many centuries, this disease was recognised as a chronic disease of the skin until the late 19th century, when its systemic manifestations were first described (Smith et al. 1988). In the modern era, this disease is one of the most common multi-system autoimmune diseases (Jacobson et al. 1997).

1.1 Epidemiology

Although individuals of all ages can be affected, this is predominantly a disease of women of child-bearing age with a female-to-male ratio of around ten to one (10:1). A community-based study during 1991 in the Birmingham and Solihull area of the West Midlands found a point prevalence of 1 in 2000 and incidence of 6.8 per 100,000 per year amongst adult women. It had been noted that it was much more common amongst women who were of Afro-Caribbean (prevalence rate 111.8/100,000, incidence 25.8/100,000/year) and South Asian descent (prevalence rate 46.7/100,000, incidence 20.7/100,000/year) as compared to Caucasian women (prevalence rate 20.7/100,000, incidence 4.3/100,000/year) (Johnson et al. 1995).

This disease has been reported worldwide but there are differences in the incidence, prevalence, pattern of organ involvement and severity of the disease among different ethnic

groups (Pons-Estel et al. 2004; Alarcon et al. 2002; Alarcon et al. 1999; McCarty et al. 1995; Serdula et al. 1979). This has been illustrated in a cross sectional study comparing SLE patients seen at three university rheumatology centres in Birmingham, Brazil and Sweden using a standard protocol in which there were differences in the severity and pattern of organ involvement among the different nationalities (Johnson et al. 1994). The incidence and prevalence of the disease appears to be on an increasing trend (Uramoto et al. 1999).

Prior to the advent of corticosteroids, this disease was considered a fatal disease with a 5-year survival rate of less than 55% (Merrell et al. 1955). Over the last few decades, the survival has improved considerably with 10-year survival rate of around 90% and this most likely reflects increased detection of milder disease and improvement in the management of the disease (Cervera et al. 2003; Abu-Shakra et al. 1995). Despite this improvement in survival, SLE patients have a much lower life expectancy than that of the general population with an average of fourfold increased risk of death and this is most pronounced in those under the age of 45 years (Moss et al. 2002). Similar mortality results were obtained from an analysis of the Birmingham SLE cohort (Leung MH, MSc thesis, University of Birmingham 2002).

1.2 Immunopathology

Multiple abnormalities of the innate and adaptive immune system are involved in the pathogenesis of SLE. It is characterised by excessive B-cells stimulation leading to autoantibody production, abnormal T-cell function, impaired clearance of immune complexes (resulting in deposition in tissues), complement activation and defective cellular apoptosis. However, these abnormalities are not uniform and there are likely to be differences between

patients and within the same patient at different time points or stages of the disease, resulting in considerable heterogeneity in its presentation and progress.

1.2.1 B-cell Abnormalities

Auto-reactive B-cells do exist in normal individuals and they produce natural autoantibodies (usually of IgM isotype) that do not cause disease or tissue damage. These natural autoantibodies do not undergo isotype switching and affinity maturation. In contrast, the autoantibodies found in SLE patients have undergone isotype switching and affinity maturation. These processes depend on cognate B-cell-T-cell interaction (Figure 1-1).

B-cells are initially activated in the T zone of the lymph node and at this site, they also present the endocytosed antigen to T-cells. This cognate interaction with T-cells is an important signal regulating isotype switching and the generation of the germinal centre response. The activated B-cells then migrate into the B cell zone to form germinal centres. They undergo rapid division and spontaneously mutate the variable region of their antibody genes in the dark zone of the germinal centre. The resultant hypermutated B-cells (centrocyte) subsequently undergo a series of processes that will determine their survival (Figure 1-2). It must be able to recognise and pinocytose surface-bound antigen from follicular dendritic cells in the light zone. This antigen is then processed and presented on its surface as part of a class II MHC molecule. A further selection step depends on cognate interaction with germinal centre T-cells, present in the outer light zone. If the antigen presented by the centrocyte is recognised by the T-cell, an interaction (via CD40 and CD40-ligand) is triggered which will override the apoptosis programme in the centrocyte and the cell then becomes either a mature memory B-cell, or a plasmablast which can differentiate into a antibody producing plasma cell, or it may recirculate within the germinal centre leading to further affinity maturation. Failure to recognise the antigen on the surface of the follicular dendritic cell or interaction

with the T-cell in the outer light zone will lead to apoptosis and deletion of the centrocyte (Salmon et al. 1999).

Figure 1-1: Development of high affinity B-cells in a germinal centre (reproduced from Salmon and Gordon 1999). B, B-cell; CB, centroblast; CC, centrocyte; FDC, follicular dendritic cells; mB, memory B-cell; PC, plasma cell; T, T-cell.

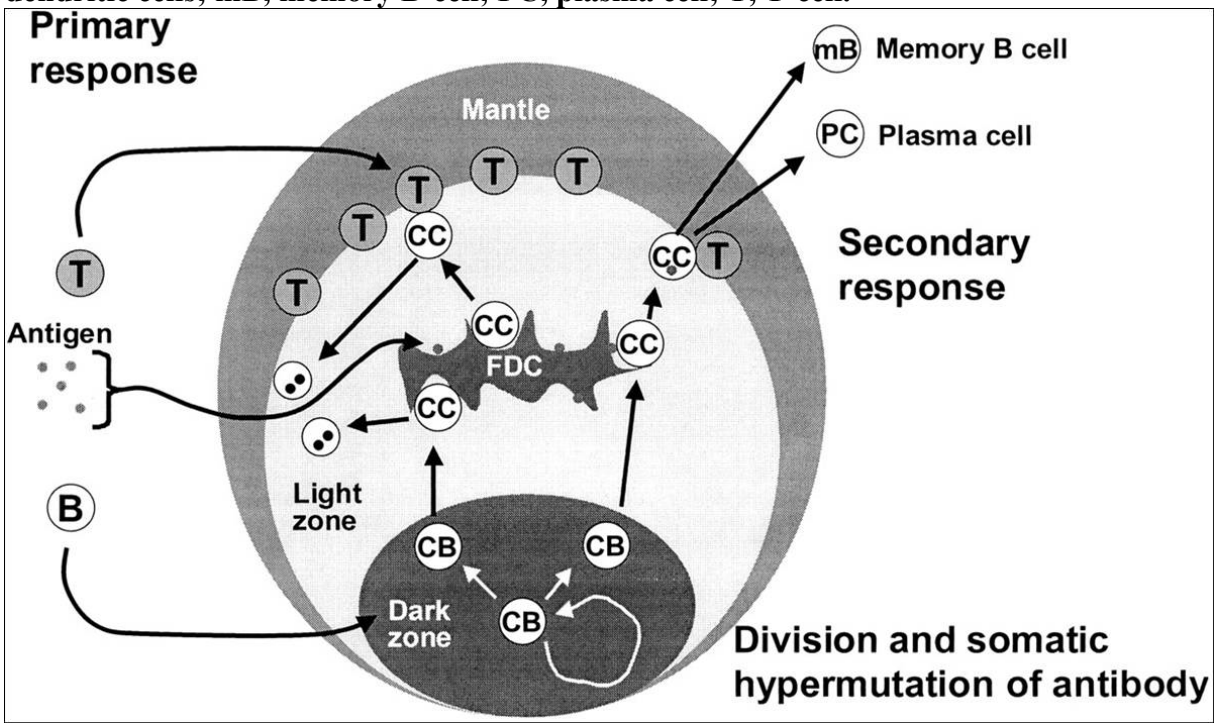
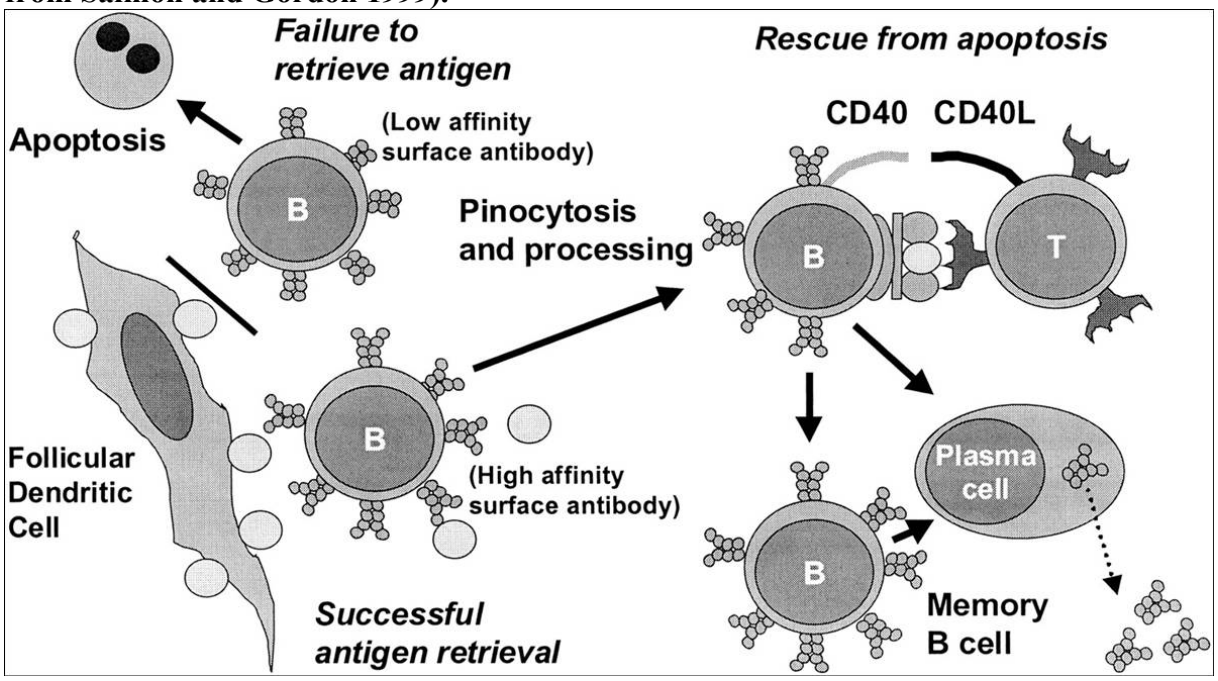


Figure 1-2: The process of selecting memory B-cells in a germinal centre (reproduced from Salmon and Gordon 1999).



In SLE, there is failure of peripheral tolerance with the presence of auto-reactive T-cells and follicular dendritic cells presenting autoantigens that allow the survival and expansion of auto-reactive B-cells leading to production of autoantibodies (Yurasov et al. 2005).

There is evidence that SLE B-cells are not just innocent bystanders producing large amounts of autoantibodies under the influence of auto-reactive T-cells, but they play a more central role in the pathophysiology of SLE. T-cell independent mechanisms of autoantibody production have been described in which DNA is able to induce immunoglobulin class switching from IgM to IgG in B-cells. This is mediated through Toll-like receptor-9 and interleukin-10, and subsequent antibody production is mediated through B-cell receptor and B-cell activating factor (BAFF) (He et al. 2004). Toll-like receptors are pattern recognition receptors that initiate the innate immune response, with toll-like receptor-9 being involved in triggering activation of B-cells.

There are intrinsic defects in SLE B-cells leading to aberrant signal transduction and exaggerated response to stimuli, resulting in B-cell hyperactivity that is a striking feature of SLE. Another contributing factor is increased production of BAFF by dendritic cells, that stimulates B-cell survival and maturation. These B-cells can activate T-cells through aberrant expression of stimulatory molecules (such as CD40L, CD80 and CD86) on the surface membrane or through its function as efficient antigen-presenting cells (Pugh-Bernard et al. 2006). In addition, activated B-cells secrete pro-inflammatory cytokines and some of these (such as interleukin-6 and interleukin-10) have a stimulatory and positive feedback effect on B-cells in an autocrine manner.

1.2.2 T-cell Abnormalities

Auto-reactive T-cells are detectable in SLE patients and also in the normal population (Hoffman et al. 1993). Normally, these T-cells have low affinity for autoantigens and are prevented from expanding by regulatory mechanisms. However in SLE, there appears to be a breakdown in immune tolerance that allows pathogenic auto-reactive T-cells to escape into the periphery and expand. T-cells that are reactive to several nuclear antigens (including DNA, histones and small nuclear ribonucleoprotein) have been found in peripheral blood of SLE patients. Upon recognition of autoantigens presented by antigen-presenting cells, these T-cells play an important role in autoantibody production by assisting class switching and affinity maturation of auto-reactive B-cells (as described in the previous section), and also produce cytokines (such as interferon-α, interferon-γ, interleukin-4 and interleukin-10) which promote autoantibody production (Hoffman 2004).

Furthermore, SLE T-cells have a hyperexcitable phenotype similar to that of B-cells. Aberrant T-cell receptors (TCR) have been found in a large proportion of SLE patients that have a lower excitation threshold than the normal TCR (Enyedy et al. 2001). In addition, there is increased quantity of lipid rafts on the T-cell membrane (Jury et al. 2004). These lipid rafts are cholesterol and gangliosides enriched membrane microdomains. They play an important role in assembly of many cell membrane-associated signalling complexes. In SLE, these rafts also appeared to be clustered in a large fraction of T-cells despite the absence of obvious stimuli. This phenotype may contribute to the heightened sensitivity of TCR signalling observed in SLE.

1.2.3 Abnormalities in Apoptosis

Apoptosis or programmed cell death is a normal physiological phenomenon which is an active, tightly regulated process leading to ordered destruction of cells. This avoids the release of intracellular contents into extracellular microenvironment that would lead to an inflammatory response. During this process, a variety of intracellular components, including nuclear components such as nucleosomes (containing DNA), Ro and La, are clustered at the surface of the apoptotic cells (Casciola-Rosen et al. 1994). These components do not normally encounter the immune system as the apoptotic cells are rapidly cleared by phagocytic cells. The complement system also helps in clearing apoptotic cells through complement receptor-mediated phagocytosis, as C1q can bind to these cells which triggers complement activation and deposition of C3 and C4 on these cells. C-reactive protein (CRP) also plays a role as it binds to apoptotic cells in a calcium-dependent manner and augments the classical pathway of complement activation, but it protects the apoptotic cells from the assembly of the terminal complement components (Gershov et al. 2000). Furthermore, CRP-enhanced opsonisation and phagocytosis of apoptotic cells by macrophages is associated with the expression of anti-inflammatory cytokine transforming growth factor β.

In SLE, there is defective clearance of apoptotic cells due to defective phagocytosis from altered maturation of phagocytes (Blanco et al. 2001). This defect is further hampered by deficiency in complement components (congenital or acquired) that is common in SLE. As a result, there is fragmentation of the apoptotic cells with release of apoptotic blebs containing intracellular antigens (autoantigens). These autoantigens can be taken up and processed by dendritic cells and presented to T and B cells, triggering the induction of autoantibodies that leads to loss of peripheral tolerance. Furthermore, these autoantigens can stimulate the immune system to produce more autoantibodies and can bind to existing autoantibodies to

form immune complexes that get deposited within tissues causing inflammatory reactions. Apart from that, autoantibodies that are present in SLE may bind to surface of apoptotic cells and the resulting antibody-mediated phagocytosis may trigger secretion of pro-inflammatory cytokines (Rovere et al. 1999). The resultant disease flare causes more apoptosis and consumption of complement components, further exacerbating the problem.

1.2.4 Autoantibodies and Immune Complexes

One of the hallmark features of SLE is the presence of autoantibodies. More than 100 autoantibodies have been described in SLE that target various autoantigens including nuclear antigens, cytoplasmic antigens and phospholipid-associated antigens (Sherer et al. 2004). The common autoantibodies in SLE, being present in more than 25% of patients, are anti-ssDNA (single-stranded DNA), anti-dsDNA (double-stranded DNA), anti-Ro, anti-poly ADP ribose polymerase, anti-histone/nucleosomes, and antiphospholipid antibodies. These autoantibodies can be broadly divided into organ-specific autoantibodies (such as anti-neuronal antibody, anti-platelet antibody) and non-organ-specific autoantibodies (such as antinuclear antibodies). The most well recognised autoantibodies associated with SLE are anti-nuclear antibodies and anti-double stranded DNA (anti-dsDNA) antibodies. Most of these autoantibodies are of IgG subtype, following isotype switching, and have the capacity to cause tissue damage.

Immune complexes are formed when autoantibodies bind to their target antigens. They may be formed within the circulation resulting in circulating immune complexes, or locally when the autoantibodies bind to antigens within tissues. Normally, immune complexes are cleared rapidly from the circulation by the mononuclear phagocyte system and this prevents tissue deposition of immune complexes that would otherwise results in inflammation and

damage. In SLE, there is impaired clearance of immune complexes by mononuclear phagocyte system (Lobatto et al. 1988).

The complement system also plays an important role as activation of the classical complement pathway by the immune complexes aids the clearance of immune complexes by phagocytes. Furthermore, activated complement components that are bound to the immune complexes inhibit immune complex interaction and prevent extension of immune complex lattice (Johnson et al. 1987). This leads to smaller immune complexes and reduced immune precipitation (that facilitates tissue deposition). This complement-mediated mechanism of immune complex clearance and prevention of immune precipitation is defective in hypocomplementaemia, which is common in SLE.

1.2.5 Complement System Abnormalities

Following the deposition of immune complexes in tissues, there is activation of the classical complement pathway and this generates pro-inflammatory fragments of complement proteins (C3a, C4a and C5a). These complement activation products attract and activate leucocytes which release various pro-inflammatory mediators resulting in inflammatory infiltrate. Activation of the complement system results in consumption of complement components leading to reduced plasma levels of classical pathway complement proteins. In addition, up to one third of SLE patients have anti-C1q autoantibodies (which are associated with lupus nephritis) that bind to activated C1 complex (bound to immune complexes) which further amplify complement activation (Walport 2002).

Paradoxically, hereditary deficiency of classical complement components is associated with a strong predisposition to developing SLE. The complement system plays an important role in maintaining immune tolerance to prevent the development of autoimmunity as it

assists with the rapid clearance of immune complexes and apoptotic cells (Walport 2002). Furthermore, it may play a role in the negative selection of auto-reactive lymphocytes that is activated through opsonisation of autoantigens by C4b or C3b (via complement receptors) (Prodeus et al. 1998). In addition, decreased serum complement C3 and C4 levels (hereditary or acquired) prevent formation of soluble immune complexes and so promotes deposition and persistence of immune complexes in tissues (Schifferli et al. 1985).

1.2.6 Cytokine Abnormalities

Cytokines have been studied extensively in human and murine SLE. A large number of cytokines have been linked to SLE immunopathology. It is beyond the scope of this section to review all the cytokines studied to date and this area has been reviewed comprehensively by Crow and Kirou (Crow et al. 2007). This section will focus on the key cytokines involved in SLE which are interferon-α, interleukin-10, interleukin-6 and tumour necrosis factor-α (TNF-α).

The high level of interferon-α expression is one of the first cytokine abnormalities described in SLE. This increased expression is further substantiated by the observation that the treatment of malignancies and hepatitis C with recombinant interferon-α occasionally results in the development of SLE (Ioannou et al. 2000). Recently, it has been found that there is a characteristic up-regulation of interferon-α gene expression, which is known as the interferon-α gene signature, in the peripheral blood mononuclear cells of SLE patients (Baechler et al. 2003). It is produced predominantly by plasmacytoid dendritic cells following stimulation with immune complexes containing apoptotic cells or nuclear fragments via Fc-gamma receptor-IIA (FcγRIIA) (Lovgren et al. 2004; Bave et al. 2003). Interferon-α induces differentiation of monocytes into efficient antigen-presenting cell capable of presenting

autoantigens, and promote survival and differentiation of activated T and B lymphocytes. When produced in abundance, this cytokine favours the development of auto-reactive immune response. The level of gene expression and serum levels of interferon-α correlated with SLE disease activity and severity (Kirou et al. 2005; Bengtsson et al. 2000).

There is increased production of interleukin-10 in SLE patients (Llorente et al. 1994) and this comes from several sources including B-cells, T-cells and monocytes. The serum level of this cytokine had been found to be correlated with SLE disease activity (Houssiau et al. 1995). It is a potent stimulator of B-cell proliferation and differentiation, thus promoting production of autoantibodies. This cytokine also has anti-inflammatory effects in which it inhibits a range of functions in both T-cells and macrophages. The effect on T-cells may have accounted for some of the T-cell abnormalities seen in SLE as some of these abnormalities were restored following treatment with anti-interleukin-10 monoclonal antibody (Llorente et al. 2000).

Interleukin-6 is a pleiotropic cytokine with wide range of effects in immune regulation, inflammation, haematopoiesis and oncogenesis. Serum interleukin-6 levels are increased in SLE and correlated with disease activity (Spronk et al. 1992). Furthermore, its production has been detected at sites of affected organs in SLE: cerebrospinal fluid (Hirohata et al. 1990), renal glomeruli (Horii et al. 1993) and skin (Nurnberg et al. 1995). It is a pro-inflammatory cytokine that promotes B-cell maturation to plasma cells and secretion of immunoglobulins. Therefore, it appears to play an important role in B-cell hyperactivity that is characteristic of SLE.

TNF-α is another pro-inflammatory cytokine with pleiotropic effects on a variety of cells. The levels are increased in SLE and appears to correlate with disease activity (Studnicka-Benke et al. 1996). The precise role this cytokine play in SLE is not clear but it is

likely to be involved in the immunopathology as TNF-α has been demonstrated within lesions of lupus nephritis (Herrera-Esparza et al. 1998) and anti-TNF-α therapy has been used in an open label study for the treatment of active disease with clinical improvement of arthritis and nephritis (Aringer et al. 2004).

1.3 Aetiopathogenesis

The aetiopathogenesis of SLE is not completely understood but is complex involving the following factors:

1. Genetic factors
2. Environmental factors
3. Hormonal factors

1.3.1 Genetic Factors

The genetic basis of SLE is supported by the ethnic differences in disease incidence and prevalence, and the tendency of this disease to aggregate in families. The degree of familial clustering, as measured by risk ratio of sibling recurrence as compared to the population, has been estimated to be as high as 29 (Alarcon-Segovia et al. 2005). Twin studies has revealed a higher concordance rate in monozygotic twins (at least 24%) than dizygotic twins (at least 2%) (Deapen et al. 1992). Therefore, SLE is considered a complex genetic trait in which there is lack of direct correlation between a phenotype and a genotype. This trait could be either due to the same genotype resulting in different phenotypes or different genotypes can result in the same phenotype. This is further complicated by the definition used to define SLE in research setting, which is dependent on fulfilling 4 of 11 American College

of Rheumatology (ACR) criteria (Hochberg 1997; Tan et al. 1982), making SLE a disease with heterogeneous phenotypes.

Genetic factors are major determinant of susceptibility to the disease and possibly contribute to the disease severity as well. The important genetic associations of SLE are:

1. Major histocompatibility complex (MHC)
2. Hereditary complement deficiency
3. Immunoglobulin receptors (Fcγ receptors)
4. Cytotoxic T lymphocyte antigen-4 (CTLA-4)
5. Programmed cell death-1 (PD-1)
6. Mannose-binding lectin (MBL)

The human MHC region is located in chromosome 6 which encodes the human leucocyte antigen (HLA) genes and many other genes that are involved in self/non-self recognition, antigen presentation and immune regulation. There has been consistent association of HLA-DR2 and HLA-DR3 with SLE in Northern European population, conferring a two-fold relative risk with each allele (Graham et al. 2002). This region encodes the MHC class II molecules that are involved in antigen presentation to T-cells.

Homozygous hereditary deficiency of the early components of the classical complement pathway is very strongly associated with development of SLE. There is a hierarchy of the prevalence and severity in the association, with C1q deficiency having the worst risk (more than 90%) and severity, followed by C4 (risk of 75%) and C2 (risk of 10%) deficiencies (Walport 2002). The C1q genes are located in chromosome 1 while the genes for C4 and C2 are within the MHC class III region of chromosome 6.

Another strong genetic association with SLE has been described for allelic variants of Fcγ receptor (FcγR) genes. FcγRs are a family of glycoprotein receptors that are expressed on the surface of leucocytes and bind to the Fc portion of IgG molecules, playing an important role in clearance of immune complexes. Three classes of these receptors have been identified (FcγRI, FcγRII and FcγRIII) and they vary in their binding capacity for IgG, their preferences for IgG isotypes, the cell types that express them and the intracellular signal they elicit. FcγRII and FcγRIII are low-affinity receptors that bind to polymeric IgG in immune complexes. Polymorphisms of FcγRIIA and FcγRIIIA have been associated with SLE and development of lupus nephritis (Karassa et al. 2003a; Karassa et al. 2003b; Norsworthy et al. 1999). These polymorphisms resulted in less efficient binding and clearance of immune complexes.

PD-1 and CTLA-4 are cell surface receptors that are structurally related, belonging to the CD28 co-stimulatory receptor family. PD-1 is expressed on activated T-cells and B-cells, while CTLA-4 is found on activated T-cells. Activation of both these receptors results in negative regulatory effects on the immune response. Polymorphisms of PD-1 (Prokunina et al. 2002) and CTLA-4 (Lee et al. 2005a) genes have been associated with SLE.

The lectin pathway leads to complement activation without the need for the presence of immunoglobulin. It is initiated when MBL binds to repetitive carbohydrate moieties (such as mannose, N-acetyl-D-glucosamine and N-acetyl-mannosamine) which are present in abundance in various micro-organisms. MBL is structurally similar to C1q and MBL variant alleles, which result in its deficiency, are associated with increased risk of SLE, in similar fashion to hereditary complement deficiency (Lee et al. 2005b).

1.3.2 Environmental Factors

Environmental factors probably play a role in triggering the autoimmune process in a genetically susceptible individual. Broadly, these factors are:

1. Ultraviolet light
2. Infections
3. Drugs
4. Heavy metals and chemicals

Ultraviolet light exposure has been well known to trigger flares of SLE and photosensitivity is a characteristic manifestation. The use of sun screen that protects against ultraviolet light has been shown to be effective in preventing ultraviolet-induced lupus erythematosus (Herzinger et al. 2004). Furthermore, ultraviolet-light exposure via a tanning device has been associated with the development of SLE in a previously healthy individual (Fruchter et al. 2005). Ultraviolet light results in apoptosis of keratinocytes and expression of autoantigens such as Ro, La, RNP and Sm on the surface of apoptotic cells (Casciola-Rosen et al. 1994). Impaired clearance of these apoptotic cells could lead to the induction of autoantibodies. In SLE patients, autoantibodies could bind to these exposed autoantigens with a resultant increase in interferon-α production and flare of disease activity (Reefman et al. 2007).

It is also a common clinical observation that infections can trigger SLE flares and may precede the development of autoantibodies and SLE. A number of infectious agents have been linked to SLE, but Epstein-Barr virus (EBV) infection is the best studied. It has been shown that more lupus patients had evidence of past EBV infection as compared to controls in both paediatric and adult age groups (James et al. 2001; James et al. 1997). The exact mechanisms

remained uncertain but many viruses induce production of proteins with epitopes that are similar to self-antigens. As a result, antibodies produced against the viral proteins could cross-react with self-antigens. These opsonised self-antigens would then be phagocytosed and presented to T-cells and B-cells resulting in loss of immune tolerance. Additionally, viruses could alter infected cells and increase their antigenicity. Apart from that, infection leads to an increase in cell death and apoptosis of cells of immune system which may drastically increase the amount of autoantigen available to activate an immune response (Salmon et al. 1999). Furthermore, viral infection induces production of interferon-α which promotes activation of the immune system and worsening of SLE disease activity.

More than 100 drugs have been implicated in causing drug-induced lupus, with procainamide and hydrallazine being the best studied. These drugs inhibit DNA methylation in T-cells and induce autoreactivity (Cornacchia et al. 1988). DNA methylation refers to the post-synthetic methylation of cytosine bases in DNA to form methylcytosine. Promoters of active genes are hypomethylated and methylation renders the genes inactive. Therefore, DNA methylation suppresses gene expression. T-cells with hypomethylated DNA over-express lymphocyte function-associated antigen 1 (LFA-1) which contributes to development of autoreactivity (Richardson et al. 1994).

Several heavy metals and chemicals have been reported to be associated with SLE, and these includes mercury, crystalline silica, gold, cadmium, vinyl chloride, hydrazines, pesticides and hair dyes. Among these, the most compelling epidemiologic data exist for crystalline silica and mercury (Cooper et al. 2004; Parks et al. 2002). However, the precise mechanisms of initiating autoimmunity by these agents remain unclear.

1.3.3 Hormonal Factors

It is not surprising that oestrogen plays a role in the expression of SLE given the strong gender bias towards the female sex in this disease and the timing of disease onset that occurs predominantly after puberty. There is also the increased risk of flare during pregnancy which is a physiological state of high oestrogen levels (Ruiz-Irastorza et al. 1996), although this has not been found uniformly in all studies. In addition, the use of oestrogen-containing contraceptive pill and postmenopausal oestrogen replacement has been associated with an increased risk of developing SLE in some studies (Sanchez-Guerrero et al. 1997; Sanchez-Guerrero et al. 1995).

Oestrogen has immunomodulatory effects as oestrogen receptors (which are nuclear receptors) are expressed in most cells of the immune system. The hormone-receptor complex binds to DNA at specific sequences along the target genes that stimulate or suppress gene transcription depending on the target cells. Oestrogen has stimulatory effects on B-cells and may rescue autoreactive B-cells from normal mechanism of tolerance, resulting in enhance production of autoantibodies. The effect of oestrogen on T-cells are less well studied but current evidence suggest it has stimulatory effects in SLE (Lang 2004).

1.4 Clinical Features

SLE is a chronic disease that is characterised by periods of exacerbations with a variable course. It is a multi-system disease that can affect any part of the body, resulting in a diverse range of clinical manifestations as shown in Table 1-1. Furthermore, involvement of a system could result in a variety of manifestations, which adds to its complexity. For example,

apart from the classical discoid rash or malar rash that are typical lupus rash, there are many other cutaneous manifestations such as bullous lesion, maculopapular rash, psoriasiform subacute cutaneous rash, annular subacute cutaneous rash, subcutaneous nodule of lupus panniculitis and cutaneous vasculitis. Similarly, involvement of the nervous system could result in many possible manifestations which include acute confusional state, psychosis, delusion, cognitive dysfunction, stroke-like syndrome, transverse myelopathy and peripheral neuropathy.

Table 1-1: Cumulative incidence of SLE manifestations (Adapted from Dubois' Lupus Erythematosus, 7th edition).

Manifestations	Cumulative Incidence (%)
Positive Antinuclear Antibodies	> 95
Fatigue/Malaise	80 – 90
Arthritis/Arthralgia/Myalgia	80 – 90
Mucocutaneous lesions	70 – 90
Positive Anti-double stranded DNA antibodies	60 – 70
Anaemia	50 – 70
Leucopenia	40 – 60
Low complements C3 and/or C4	40 – 60
Pleurisy/Pericarditis	40 – 60
Fever	50
Cognitive dysfunction	50
Renal involvement	40 – 60
Hypergammaglobulinaemia	30 – 40
Anti-Sm antibody	10 – 30
Thrombocytopenia	20 – 30
Psychosis	10 – 30
Seizures	10 – 20

Recently, involvement of the gastrointestinal tract (such as intestinal vasculitis, pancreatitis, cholecystitis, hepatitis, abdominal serositis, intestinal pseudo-obstruction and

malabsorption) and eye (such as keratitis, scleritis, uveitis and vaso-occlusive retinopathy) have become increasingly recognised (Sivaraj et al. 2007; Sultan et al. 1999). Although these manifestations are uncommon, it is critical to recognise them early to enable early institution of immunosuppressive therapy, as they carry significant morbidity and mortality.

A hallmark feature of this disease is the production of variety of autoantibodies. The most well recognised autoantibodies associated with SLE are anti-nuclear antibodies and anti-dsDNA antibodies. Some of these antibodies are specific to SLE such as anti-dsDNA and anti-Sm, hence are useful as diagnostic markers. The other characteristic immunological manifestations are the formation of immune complexes and activation/consumption of complement.

As there are myriad of possible manifestations with SLE and most of these manifestations are not specific to the disease itself, it is not surprising that the clinical presentation of this disease could mimic other medical or surgical conditions. Therefore, it is not uncommon for it to be confused with other conditions such as infection and fibromyalgia as the clinical presentation can be similar. Furthermore, as any organ system may be involved, this disease may present in variable combinations of organ system manifestations that can vary between patients and within the same patient over time. All this can lead to a delay in the diagnosis of the disease or identification of a flare in the disease with potentially disastrous consequences.

The American College of Rheumatology (ACR) has devised a classification criteria for SLE that was developed for clinical trials and research studies (Table 1-2) (Hochberg 1997; Tan et al. 1982). There are 11 criteria and any 4 or more criteria are required for the classification of SLE. However, it should be noted that this classification criteria were not intended for diagnostic purposes. This is demonstrated in a study which revealed that over

50% of patients with SLE did not fulfil the criteria at a particular point in time, although the majority eventually did after several years (Levin et al. 1984). Apart from that, the criteria set is heavily weighted towards mucocutaneous manifestations and there are some important omissions such as kidney biopsy and activation of complements (Petri et al. 2004a). The classification criteria are currently in the process of being revised, to take into account these concerns.

Table 1-2: The 1997 ACR revised criteria for classification of SLE.

Criterion	Definition
Malar rash	Fixed malar erythema, flat or raised
Discoid rash	Erythematous-raised patches with keratotic scaling and follicular plugging; atrophic scar may occur in older lesions
Photosensitivity	Skin rash as an unusual reaction to sunlight, by patient history or physician observation
Oral ulcers	Oral or nasopharyngeal ulcers, usually painless, observed by physician
Arthritis	Non-erosive arthritis involving two or more peripheral joints, characterised by tenderness, swelling or effusion
Serositis	1. Pleurisy (convincing history of pleuritic pain or rub heard by physician or evidence of pleural effusion) or 2. Pericarditis (documented by ECG, rub or evidence of pericardial effusion)
Renal disorder	1. Persistent proteinuria (>0.5 g/day or >3+) or 2. Cellular casts of any type
Neurological disorder	1. Seizures (in absence of other causes) or 2. Psychosis (in absence of other causes)
Haematological disorder	1. Haemolytic anaemia or 2. Leucopaenia (< 4000/ml on two or more occasions)

	or 3. Lymphopaenia (<1500/ml on two or more occasions)
Immunological disorder	1. Anti-double-stranded DNA or 2. Anti-Sm or 3. Positive finding of antiphospholipid antibodies based on (a) abnormal serum level of IgG or IgM anticardiolipin antibodies, (b) positive test for lupus anticoagulant using a standard method, or (c) a false positive serological test for syphilis known to be positive for at least 6 months
Antinuclear Antibodies	An abnormal titre of antinuclear antibody by immunofluorescence or an equivalent assay at any time and in the absence of drugs known to be associated with "drug-induced lupus syndrome"

1.5 Outcomes in SLE

Outcomes in SLE can be described by three domains, namely disease activity, accumulated damage and health status of patients (Isenberg et al. 1999). Disease activity is defined as immune-mediated disease process that is potentially reversible as opposed to damage which is an irreversible process or 'scarring'. The assessment of disease activity will be discussed further in Section 1.7.

Mortality is the ultimate form of damage from the disease and remains an important outcome measure that is used to inform the management of patients with SLE. With improvement in management of SLE, the survival of patients with the disease has improved substantially. As such, other measures of damage have become necessary to complement the mortality statistic in this chronic disease. A validated scale has been developed for this purpose and is known as Systemic Lupus International Collaborating Clinics/American College of Rheumatology damage index (SLICC/ACR DI) (Gladman et al. 2000a; Gladman

et al. 1997; Gladman et al. 1996). This index looks at damage that has developed since the diagnosis of SLE and this is regardless of attribution, whether it is due to disease activity, its therapy or intercurrent illness. It divides the items of damage into 12 organ systems (Ocular, Neuropsychiatric, Renal, Pulmonary, Cardiac, Peripheral Vascular, Gastrointestinal, Musculoskeletal, Skin, Gonadal Failure, Diabetes Mellitus and Malignancy). Its score ranges from 0 indicating no damage up to a possible maximum of 47 and this accumulates over time. A multi-national study of more than 1000 patients, who were followed-up longitudinally, showed that higher SLICC/ACR DI scores were at higher risk of death (Gladman et al. 2000a). In another study, early renal damage (within 1 year of diagnosis) as assessed using this index was predictive of development of end-stage renal failure while pulmonary damage was predictive of mortality at 10 years after diagnosis (Stoll et al. 1996).

Health status of the patient or health-related quality of life (HRQOL) is the person's sense of physical, emotional and social well-being associated with the disease or its treatment. This provides the patient's perspective of the impact of the disease. It is usually assessed using a self-completed questionnaire. HRQOL can be measured using generic instruments that are applicable to various diseases, such as Medical Outcomes Study Short Form 36 (SF-36) and European QoL. However, these generic instruments may not capture certain aspects (such as sleep disturbance, sexual functioning, body image and impact on the family) that are relevant to SLE and hence may not be sufficiently responsive in clinical trials. As a result, HRQOL instruments that are specifically designed for SLE, have been developed recently such as SLEQOL (Leong et al. 2005) and LupusQOL (McElhone et al. 2007).

1.6 Treatment

The treatment of SLE is directed towards the control of disease activity, prevention of development of damage and management of complications. This section is not intended to be a comprehensive review on the treatment of SLE. The focus of this section is on the principles of treatment to control disease activity, thereby preventing development of long term complications and mortality from uncontrolled immune-mediated tissue damage. There are numerous options available for the treatment of SLE disease activity, which can be categorised into:

1. Symptomatic
2. Antimalarials
3. Prasterone
4. Thalidomide
5. Dapsone
6. Retinoids
7. Corticosteroids
8. Immunosuppressives/Cytotoxics
9. Immunoglobulins
10. Plasmapheresis
11. Rituximab

Treatment is tailored to the patient and the major determinant of therapeutic option used is the level of disease activity that can be divided broadly into mild, moderate and severe. Severe disease activity is defined as manifestations that are life threatening or cause significant organ dysfunction such as seizures, nephrotic syndrome or inability to perform activities of daily living due to inflammatory arthritis. In contrast, mild disease activity refers

to minor manifestations that only cause some discomfort and the patient is able to continue with his/her daily routines.

Severe disease activity requires aggressive therapy with high dose corticosteroids (more than 20 mg/day of prednisolone or equivalent) and usually together with an immunosuppressives. Thus far, clinical trials in severe disease activity have been for lupus nephritis and have shown efficacy of cyclophosphamide (Austin, III et al. 1986), mycophenolate mofetil (Ginzler et al. 2005) and intravenous methylprednisolone (in combination with intravenous cyclophosphamide) (Illei et al. 2001). However, the efficacy in these studies has been defined based on renal outcome only (such as reduction in proteinuria and reduction in worsening of renal function). Intravenous immunoglobulins and plasmapheresis are mainly used in those with life-threatening complications. Recently, rituximab which is a chimeric monoclonal antibody against CD20 that causes depletion of B-cells, has shown promise in the treatment of severe disease activity (Isenberg 2006).

Moderate disease activity is usually treated with lower doses of corticosteroids (up to 20 mg/day prednisolone or equivalent). In addition, immunosuppressives are commonly used concomitantly with corticosteroids. Antimalarials have been shown to be safe and effective with mucocutaneous manifestations and inflammatory arthritis, hence can be used instead of immunosuppressives for these manifestations (D'Cruz 2001). For certain manifestations, local treatment with corticosteroids (and occasionally immunosuppressives) could be used such as intra-articular injection of methylprednisolone for inflammatory synovitis or topical corticosteroids for discoid rash. Thalidomide, retinoids and dapsone are reserved for treatment of refractory lupus rash. Prasterone is an androgen which has been shown recently to have steroid-sparing properties and to reduce the number of SLE flares (Petri et al. 2004b; Petri et al. 2002).

Mild disease activity usually only requires symptomatic treatment such as analgesia or non-steroidal anti-inflammatory drugs (NSAIDs). Not uncommonly, antimalarials or topical corticosteroids, with their good safety profile, are also used for milder mucocutaneous or musculoskeletal manifestations that are recurrent or persistent.

Despite the availability of numerous therapeutic options, it is surprising that there has not been a new treatment approved for use in SLE by the drug authorities in United States of America and Europe for the past 20 years. This has been due to the scarcity of randomised placebo-controlled trials that are adequately powered. One of the main reasons for this is the difficulty in designing and conducting clinical trials in SLE (Mukhtyar et al. 2007; Dall'Era et al. 2006). This reflects the complexity and multi-system nature of this disease which makes objective assessment of disease activity difficult and challenging. With the development of several promising new agents, there is a pressing need for this issue to be addressed quickly in order to assess these therapies with the greatest likelihood of success.

1.7 Assessment of Disease Activity

As with most human attributes, SLE disease activity cannot be measured directly. What is being measured is the observed phenomenon that allows us to infer the existence of these attributes. Standardised and objective assessment of disease activity is crucial as it allows for comparison of results of studies between different centres and enables multicentre studies to be conducted (Mukhtyar et al. 2007). This is essential in clinical trials to determine the efficacy of any particular treatment and to compare different therapeutic approaches.

Before a measure can be used as a marker of disease activity, it needs to undergo rigorous validation process to ensure that it is biologically relevant, reliable, reproducible and sensitive to change (Singh et al. 2006; Illei et al. 2004a). This is imperative as any measure

that lacks validity is measuring a characteristic (in whole or part) other than that which is intended. In general, these markers can be in the form of biological markers (biomarkers) or composite clinical disease activity indices.

1.7.1 Biomarkers

A biomarker is defined as a genetic, biological, biochemical or molecular characteristic that reflects a biological process (which in this context is disease activity), and can be evaluated qualitatively or quantitatively. With numerous abnormalities of the immune system having been identified, numerous candidate markers have been proposed as biomarkers of disease activity. The most widely used classical markers of disease activity are the serum anti-dsDNA antibodies and complement C3 and C4 levels.

There is a divergent view on the usefulness of anti-dsDNA antibodies and complement C3 and C4 levels in predicting SLE disease activity with some studies showing good association between these classical markers and disease activity, while other studies indicate that these classical markers are of little value. This difference in opinion is mainly the result of difference in study design used, difference in the type of organ involvement in recruited patients and different definition of disease activity being employed in the studies. Two Dutch longitudinal studies (Ter Borg et al. 1990; Swaak et al. 1986; Swaak et al. 1982) in which anti-dsDNA and complement C3 and C4 levels in patients were measured at close intervals (4 to 6 weekly) demonstrated that exacerbation in disease activity were preceded by rapid increase in anti-dsDNA levels (over a mean period of 8 to 10 weeks), peaking at the time of exacerbation and this was followed by rapid reduction in the levels. A more gradual reduction in complement C3 and C4 levels (over a period of up to 24 weeks) was also noted to precede the flares. The association with flare was stronger for anti-dsDNA antibodies compared to

complement levels. Patients with stable levels of anti-dsDNA antibodies were much less likely to have a flare of disease activity. In these two studies, the flares were predominantly renal involvement and it was noted that all renal flares were preceded by increase in anti-dsDNA levels. The definition of flare used in these two studies were well defined but unfortunately had not been validated (Ter Borg et al. 1990; Swaak et al. 1986; Swaak et al. 1982).

Similar results were obtained in another prospective longitudinal study in which patients were assessed monthly and disease activity was measured using several composite clinical activity indices and physician's global assessment (Ho et al. 2001b). However, decreasing complement levels were only predictive of renal and haematological flares (Ho et al. 2001a). In a study by Esdaile et al, changes in anti-dsDNA antibodies and complement levels were not found to be good predictors of flare (Esdaile et al. 1996). However, anti-dsDNA antibodies and complement levels were measured at the time of flare and around 3, 6 and 9 months prior. The definition of flare was an increase of at least 6 in the modified SLEDAI score (after excluding 9 items), which was scored retrospectively. This study also highlighted the difference in the results between cross sectional and longitudinal evaluation of these markers. Cross sectional study of these markers found that patients with renal flare tended to have higher anti-dsDNA levels and lower C3 levels, in contrast to longitudinal assessments where there was no association between these markers and flare.

Despite the differences in the results, what can be concluded from these studies is that longitudinal changes in anti-dsDNA and complement levels are probably more helpful than a single estimation in predicting flares. Rapid increase in anti-dsDNA levels is useful in predicting flares of disease activity in some but not all patients. Therefore, it is of greatest use when it is measured frequently (every 4 to 6 weeks). Decreasing complement C3 and C4

levels are also predictive of flares but its association is weaker when compared to anti-dsDNA and the decrease is much more gradual. It has been consistently shown that increasing anti-dsDNA and decreasing complement levels are strongly associated with lupus nephritis.

As serum complement levels represent the dynamic state of complement synthesis and consumption, the levels may not reflect the true status of complement activation which is a feature of active disease. Firstly, there is a wide variation in normal serum complement levels among different individuals (partly due to genetic polymorphisms). Secondly, there is a control in the balance of synthesis and consumption of complement proteins whereby increased consumption is counterbalanced by an increase in synthesis that is variable between individuals. Additionally, many of the complement proteins are acute phase reactants and as such, there will be an increase in the synthetic rate in response to inflammatory stimuli. The drawbacks of serum complement levels led to the rationale that plasma levels of complement activation products (such as C3a, C4a, C5a, C3d, C4d, C_{5b-9}, Ba and Bb) would be better marker of disease activity. Complement split products, particularly of the alternative and terminal complement pathway activation, had been shown to be more useful in predicting flare of disease activity than C3 and C4 levels (Buyon et al. 1992). However, these assays have not been widely adopted outside of a research setting as accurate measurement of complement activation products is hampered by the instability and short half-lives of these activation products, whereby complement activation can easily occur in vitro after blood sampling. Apart from that, complement activation is not specific to SLE disease activity but also occurs in other inflammatory conditions such as infection (Sturfelt et al. 2005). This may have contributed to the inconsistent results on the utility of complement split products with disease activity from other studies (Mollnes et al. 1999).

Numerous cytokines have been studied with regards to its association with disease activity, which has been reviewed thoroughly by Ilei et al (Illei et al. 2004b). Data on a few cytokines have shown promise, namely interferon-α, soluble interleukin-2 receptor and soluble TNF receptors, but further studies (particularly with larger number of patients and in a longitudinal fashion) are required to confirm the utility of these cytokines as markers of disease activity. One of the major drawbacks of cytokines as markers of disease activity is that they have pleiotropic effects with stimulatory and inhibitory effects on different cells of the immune system. Furthermore, the production of these cytokines is context dependent varying with many factors such as stage of the disease (advanced versus early), presence of organ damage, type of organ involvement, treatment status and ethnicity. This does limit the usefulness of cytokines as reliable markers of disease activity.

Other potential markers of disease activity which require further confirmatory studies include erythrocyte or reticulocyte bound C4d (Liu et al. 2005; Manzi et al. 2004) and circulating $CD27^{high}$ plasma cells (Jacobi et al. 2003)

Although a large number of studies have described potential markers of disease activity, none of them has fulfilled the criteria of a validated biomarker to date. The majority of these studies were not designed to validate the marker as a biomarker. Many of the studies are cross sectional, which is inadequate as longitudinal studies are required to demonstrate that the marker can be used to monitor disease activity in individual patients. It is critical that the marker is studied across different patient population characteristics, such as ethnicity, organ involvement, treatment status and stage of disease (early versus late). The definitions of active disease used have not been uniform across studies to date and standardised assessment of disease activity was not used in many studies, making comparison of results difficult and probably contributed to the conflicting results. A large number of studies also lack the

statistical rigor necessary for valid conclusions. Last, but not least, many of the bioassays used to measure the marker are not standardised resulting in conflicting results from different laboratories (Illei et al. 2004a).

1.7.2 Composite Clinical Disease Activity Indices

Composite clinical disease activity indices are instruments that combine clinical and laboratory parameters in the evaluation of SLE disease activity. Numerous composite clinical disease activity indices have been developed to standardise the assessment of disease activity in SLE. There are over 60 different disease activity indices that have been described in the literature, but most of them have not been validated (Liang et al. 1988). Of these, there are 5 disease activity indices that have undergone validation:

1. SLE Disease Activity Index (SLEDAI) (Bombardier et al. 1992)
2. Systemic Lupus Activity Measure (SLAM) (Liang et al. 1989)
3. European Consensus Lupus Activity Measurement (ECLAM) (Vitali et al. 1992)
4. Lupus Activity Index (LAI) (Petri et al. 1992)
5. Classic British Isles Lupus Assessment Group Index (Classic BILAG) (Hay et al. 1993)

Broadly, there are two types of disease activity indices. The majority of currently available indices are global score indices in which the scores of the individual items within the index are summated into a total score that provides an overall measure of disease activity. In contrast, individual organ-system score index provides scores for each individual organ-system within the index reflecting the level of disease activity in these organ-systems. The only individual organ-system score index is the Classic BILAG index. Currently, the two

indices commonly used in studies and clinical trials are the SLEDAI and Classic BILAG indices, which will be discussed further.

SLE Disease Activity Index

This global score index was developed in 1985 from a meeting in Toronto. The 24 items (16 of which were clinical) to be included in the index were identified through nominal consensus process involving 15 lupus experts, who completed a questionnaire containing a list of 37 clinical variables and rated these variables on the importance in measuring disease activity. Using 574 paper patient profiles, weighting for each item was determined by multiple regression analysis. The weighting used ranged from 1 to 8, giving rise to a total possible score of 105 (Bombardier et al. 1992). A manifestation is recorded if it has been present at any time over the past 10 days regardless of severity, or whether it has improved or worsened.

This index had undergone some form of validation. It was shown to be reliable by several studies (Hawker et al. 1993; Guzman et al. 1992; Petri et al. 1992; Gladman et al. 1992; Liang et al. 1989), however these studies suffered from small number of patients (less than 40) and incorrect use of simple correlation to assess reliability in one study (Guzman et al. 1992). Construct validity of the index was demonstrated as it had good correlation with physician's global assessment and other activity indices (Ward et al. 2000; Gladman et al. 1994; Guzman et al. 1992; Petri et al. 1992; Gladman et al. 1992; Liang et al. 1989). Once again, all the studies except one involved small number of patients (less than 50 patients) and paper patient profiles were used in one study (Gladman et al. 1994). Its sensitivity to change was demonstrated by several studies which used internal responsiveness methods (discussed in section 1.8.4) especially in the form of effect size or standardised response mean (Chang et al. 2002a; Fortin et al. 2000; Ward et al. 2000; Brunner et al. 1999; Guzman et al. 1992; Liang

et al. 1989). Additionally, Fortin et al used the more robust external responsiveness method comparing changes in the index with the corresponding change in physician's opinion of change in patient's disease status (Fortin et al. 2000). These sensitivity to change studies also involved relatively small number of patients (less than 100 patients)

It was subsequently noted that the original index focused on new or recurrent manifestations and failed to capture on-going activity. This led to a revision by the Toronto group which allowed the index to capture on-going disease activity in the items of proteinuria, rash, alopecia and mucosal ulcers, giving rise to SLEDAI-2000 index (Appendix 3) (Gladman et al. 2002). Another group of investigators in the United States, who was involved in the trial of oestrogen in SLE, produced another version known as the SELENA-SLEDAI that incorporated on-going disease activity in the items rash, alopecia and mucosal ulcers, but also made changes to other descriptors such as including vertigo in cranial nerve disorder, changing the criteria for proteinuria to include a recent increase of more than 0.5g/day, and allowing only one of the features in the descriptor for pleurisy and pericarditis to be present for it to be recorded (instead of the combination of features required in the original index such as pleuritic chest pain with pleural rub) (Petri et al. 2005). Unfortunately, these two revisions have not been formally validated. SLEDAI-2000 had been shown to correlate very well with the original SLEDAI index (Gladman et al. 2002), but this was to be expected as SLEDAI-2000 index is derived from the original SLEDAI index and the majority of the items were identical.

Classic British Isles Lupus Assessment Group Index (Classic BILAG)

The original BILAG index was developed in 1984 through a nominal consensus approach by the British Isles Lupus Assessment Group (BILAG) which was a group of

rheumatologists with a special interest in the care of SLE patients (Symmons et al. 1988). It was developed based on the principle of the physician's intention to treat, on the premise that there was broad agreement among clinicians as to which aspect of the disease required treatment with disease-modifying therapy such as corticosteroids or immunosuppressives. This first version contained 87 items distributed over 8 systems (Non-specific, Mucocutaneous, Neurological, Renal, Musculoskeletal, Cardiorespiratory, Vasculitis and Haematological). It had an ordinal scoring scale by design and did not have a global score. Disease activity was classified into 4 categories from Grade A to D:

Grade A – active disease requiring immediate treatment with disease-modifying therapy

Grade B – active disease but less severe than A requiring symptomatic treatment

Grade C – stable disease

Grade D – inactive disease

Several issues with the first version were realised such as poor reliability in the musculoskeletal and vasculitis systems, inappropriate scoring of Grade B in neurological system with migraine headaches, and the inability of the index to differentiate between patients with mild stable involvement and those with past involvement of a system. This led to revisions of the index and the third revision (version 3 which is now known as the Classic BILAG index) was subsequently validated and published (Appendix 2) (Hay et al. 1993). This revision involved the following modifications: definition of a time scale, provision of a glossary, changes to the scoring scheme and refinement of the definition of items to avoid ambiguity. It has 86 items distributed over 8 systems (General, Mucocutaneous, Neurology, Musculoskeletal, Cardiorespiratory, Vasculitis, Renal and Haematology) that records manifestations occurring over the past month as compared to the month before. As such, this is a transitional index that is able to capture changing severity of clinical manifestations as

most items in the index (except for laboratory investigations and a few items that are yes/no) are recorded, if present, as new, improved, the same or worse over the past month as compared to the previous month (see Appendix 2). The scoring of this index had been modified so that there were 5 categories of disease activity from A to E:

Grade A – very active disease requiring immunosuppressive drugs and/or prednisolone dose of more than 20 mg daily (or equivalent)

Grade B – moderate disease activity requiring lower dose of corticosteroids, antimalarials or NSAIDs

Grade C – stable mild disease

Grade D – no current disease activity but the system has previously been affected

Grade E – no current or previous disease activity in the system

The glossary was further expanded in 1999 and minor modifications were made to the scoring of renal and neurological systems (Isenberg et al. 2000). Even though this index was developed based on the principle of intention to treat, it was devised to capture manifestations of disease activity and the treatment has no bearing on the scoring of this index.

This index was shown to be reliable in a multi-centre study involving 82 patients in which succinct case summary was provided (Hay et al. 1993). The same investigators also demonstrated the construct and criterion validity of the index. Criterion validity was assessed in a large multi-centre study of 353 patients looking at the agreement between Grade A score and commencement or increase in prednisolone (dose of more than 20mg daily) or immunosuppressives. However, construct validity was only analysed using data from one of the centres that recruited 127 patients and the association between Grade A score with ESR and anti-dsDNA antibodies was assessed. The sensitivity to change of this index was shown

in two small longitudinal studies (less than 50 patients each) using internal responsiveness methods (Ward et al. 2000; Brunner et al. 1999).

1.7.3 Problems with Current Indices

Despite the availability of numerous disease activity indices, there are key problems with them. Apart from the Classic BILAG index, the others indices are not transitional in nature. A transitional index is one that allows the change in severity of manifestations to be recorded, hence improvement or worsening of manifestations are captured. As a result, in a non-transitional index that is not able to differentiate change in severity of manifestations, a manifestation that has improved considerably will derive the same score as the same manifestation that has worsened significantly. This leads to incorrect scoring with over-scoring for manifestation that has improved and under-scoring for manifestation that has worsened, and potentially the wrong conclusions could be drawn from such an analysis. It is an essential requirement for an index to be transitional before it can be used to assess the efficacy of therapies.

Global score which is a feature of most indices could be misleading. Firstly, the definitions of different levels of activity (mild, moderate or severe) based on the global score has not been established with certainty for global score indices. This is clearly an important issue as cut-off scores are used in clinical studies to stratify patients into different levels of disease activity and also to determine eligibility for studies, especially in a clinical trial. There have been only two studies that have assessed the best cut-off score to define active disease for SLEDAI and SLAM indices (Gladman et al. 2000b; Abrahamowicz et al. 1998). Even these two studies differ in the results due to methodological differences. The study by Abrahamowicz et al used a small sample (n = 30) of abstracted case histories that were

assessed by 38 lupus experts and suggested that the best cut-off score based on treatment decision for SLEDAI was 6 and for SLAM was 7. The other study by Gladman et al used a larger database of 230 patients and based on physician's global assessment of disease activity, determined that the best cut-off score for SLEDAI was 4. However, the physician's assessment of disease activity was done retrospectively from the records and only descriptive statistics were used in the analysis with no allowance made for possible correlation between repeated observations from the same patient. In addition, the physician's global assessment of disease activity (PGA) is not ideal as the gold standard for disease activity, as several studies have shown that PGA is associated with unsatisfactory performance and poor agreement between expert physicians (Liang et al. 2004; Wollaston et al. 2004; Brunner et al. 1999; Gladman et al. 1994; Gladman et al. 1992).

Secondly, there are many different permutations that can lead to the same global score. To illustrate this, a score of 4 in SLEDAI-2000 can be achieved with arthritis, proteinuria or a combination of rash, fever and thrombocytopenia, as the items have different weighted scores (see Appendix 3 for details of item weighting). Hence, the same global score over two time points does not necessarily mean that the disease activity has remained unchanged. Furthermore, the minimal clinically important change for these global score indices has yet to be established. This is another important issue as it is required to classify whether the patient's disease activity over time has improved, worsened or remained unchanged. There have been a few studies that have tried to address this issue using different methods, but unfortunately they were hampered by small sample size (less than 100 patients) (Fortin et al. 2000), use of simple descriptive statistics for analysis (Gladman et al. 2000b), use of abstracted case history (Liang et al. 2004) or use of physician's global assessment as the gold standard for disease activity (Liang et al. 2004; Fortin et al. 2000; Gladman et al. 2000b; Petri

et al. 1991). Hence, it is not surprising that all these studies yielded different results with the same index as illustrated by the minimal increase in SLEDAI score associated with a worsening of disease activity ranging from 3 to 8 across these studies. This could lead to problems with misclassification of response and difficulties in the interpretation and analysis of results. This may have contributed to the high placebo response seen in the clinical trial of prasterone in which SLEDAI was used to assess disease activity (Petri et al. 2004b; Petri et al. 2002).

Many of these indices are not comprehensive enough as they are unable to capture certain manifestations of active disease that are not included in the index. This is by design as they are meant to be easy to use and as a result, only relatively common and/or severe manifestations are incorporated into the indices. Unfortunately, there is a major drawback to this design due to the resultant misclassification, as patients with manifestations that are not captured by the index are wrongly classified as having inactive or less active disease than they are actually suffering from. For example, serious SLE manifestations such as myelopathy or peripheral neuropathy would not be captured by SLEDAI-2000. Although the Classic BILAG index has been designed to be a comprehensive index, it is not immune to such criticism as it does not capture most gastrointestinal or ophthalmic manifestations of SLE, which are increasingly recognised.

Apart from these 5 disease activity indices (Classic BILAG, SLEDAI, LAI, SLAM and ECLAM), none of the other reported indices have been validated. Although these 5 indices have been generally accepted as 'validated', they have not undergone the comprehensive and rigorous validation process that is required of an index before it can be said to be truly valid. Many of the validation studies suffered from small sample size, use of paper patients, retrospective assessments and inadequate statistical analysis. In addition, the

sensitivity to change of these indices has not been adequately assessed. Most of the studies have used an internal responsiveness method to demonstrate that these indices are sensitive to change (Chang et al. 2002a; Fortin et al. 2000; Ward et al. 2000; Brunner et al. 1999; Guzman et al. 1992; Liang et al. 1989). This method only shows that these indices do change with time, but this has not been compared to an external standard. Therefore, these changes may not be clinically meaningful. It is important that the change in the index over time reflects actual change in the clinical status of the patient when the index is used to measure transition in clinical states. One study did use an external responsiveness method to assess sensitivity to change, comparing the change in the index against the change in an external standard (Fortin et al. 2000) but the external standard used was physician's global assessment which had been shown to be unsuitable for this purpose (Liang et al. 2004; Wollaston et al. 2004; Brunner et al. 1999; Gladman et al. 1994; Gladman et al. 1992).

Therefore, the currently available disease activity indices cannot be considered adequate for capturing changes in disease activity and require substantial improvement as many of them are clearly not good enough for use in a clinical trial setting.

1.8 Development and Validation of the BILAG-2004 Index

Over the past few years, concerns regarding the Classic BILAG index were raised amongst members of BILAG:

1. It became apparent that this index had some serious omissions, particularly with respect to gastrointestinal and ophthalmic manifestations. These have not been well studied previously but recent studies suggest that they are more common than previously thought and associated with significant morbidity and mortality (Sivaraj et al. 2007; Lee et al. 2002; Hallegua et al. 2000).

2. The glossary for the index was far too brief and inadequate. Furthermore, much of the terminology used was outdated, especially in the neuropsychiatric system where the definitions predate the development of the American College of Rheumatology (ACR) nomenclature and case definitions for neuropsychiatric lupus syndromes (ACR Ad Hoc Committee on neuropsychiatric lupus nomenclature 1999).
3. Our understanding and perception of the significance of certain features of the disease have changed considerably over the last 20 years. For certain severe manifestations such as myelopathy, it is now accepted that thrombosis is contributory in addition to the inflammatory process and it is not uncommon for these manifestations to be treated with anticoagulation (in combination with immunosuppressives and/or systemic corticosteroids).
4. Some of the items are in fact damage items and should not be included, such as avascular necrosis, sclerodactyly, telangiectasia and tendon contractures.
5. Improvement of items scoring Grade A resulted in a decrease in the score to Grade C which is inappropriate for such serious manifestations that are still on-going.

Two members of the BILAG (Caroline Gordon and David Isenberg) proposed in 2002 that the Classic BILAG index needed to be revised, taking into account the above-mentioned issues. Conceptual change to the index started in 2003 with all members of BILAG being engaged in intensive consultation/discussion about how best to revise the index and the preliminary new index was completed in 2004, hence the name 'BILAG-2004 index'.

The BILAG-2004 index is based on the Classic BILAG index but with the following changes incorporated:

1. Addition of gastrointestinal and ophthalmic manifestations

2. The terminology, glossary and definitions have been updated and improved to reflect current understanding of the disease. This includes the incorporation of the ACR neuropsychiatric nomenclature/definitions and major revision to the majority of items in the other systems.
3. The scoring scheme has been refined and this includes recognition that anticoagulation (in the presence of intensive immunosuppression) may be used to treat active manifestations and that improvement in items that score Grade A results in a Grade B score (instead of Grade C).
4. Removal of items that are clearly not due to disease activity such as avascular necrosis, tendon contractures, thromboembolism, telangiectasia, calcinosis, sclerodactyly, Raynaud's phenomenon, livedo reticularis and superficial phlebitis.
5. Redistribution of items in the Vasculitis system into individual systems.

Like its predecessor, the BILAG-2004 index is a transitional index but has more items (102 in total) distributed over 9 systems (Constitutional, Mucocutaneous, Neuropsychiatric, Musculoskeletal, Cardiorespiratory, Gastrointestinal, Ophthalmic, Renal and Haematological). The categorical (ordinal) scoring system is similar to that of the Classic BILAG index but with some modification to include an expanded list of treatments that were considered appropriate for a certain level of disease activity, as follows:

Grade A – very active disease requiring any of the following:

1. systemic high dose oral corticosteroids (equivalent to prednisolone > 20 mg/day)
2. intravenous pulse corticosteroids (equivalent to pulse methylprednisolone ≥ 500 mg)

3. systemic immunomodulators (including biologicals, immunoglobulins and plasmapheresis)
4. therapeutic high dose anticoagulation in the presence of high dose steroids or immunomodulators (such as warfarin with target INR of 3 to 4)

Grade B – Moderate disease activity requiring any of the following treatment:

1. systemic low dose oral corticosteroids (equivalent to prednisolone ≤ 20 mg/day)
2. intramuscular or intra-articular or soft tissue corticosteroids injection (equivalent to methylprednisolone < 500mg)
3. topical corticosteroids
4. topical immunomodulators
5. antimalarials or thalidomide or prasterone or acitretin
6. symptomatic therapy (such as NSAIDs for inflammatory arthritis)

Grade C – mild disease

Grade D – no current disease activity but the system has previously been affected

Grade E – no current or previous disease activity in the system

Before a newly developed index can be used widely, it needs to be validated to ensure that it is measuring what it is purported to measure. The validation process can be divided into the following processes:

1. Face and content validity
2. Reliability
3. Construct and criterion validity
4. Sensitivity to change
5. Predictive validity

6. Rasch analysis

The preliminary BILAG-2004 index had undergone initial validation with a reliability study involving 8 patients being assessed by 8 physicians, before work for this thesis began (Isenberg et al. 2005). Two reliability exercises were performed and there was good reliability for most systems in the index. There was some concern regarding the performance of the musculoskeletal system that was felt to be most likely related to inadequacies in the glossary. Further changes were made to improve the glossary after the reliability study (Appendix 1).

1.8.1 Face and Content Validity

Face validity and content validity have overlapping concepts and they go hand in hand, hence they are commonly assessed together. Face validity refers to the credibility of the index and it assesses whether the index appears reasonable or sensible. Content validity on the other hand determines the comprehensiveness of the index and assesses if the index covers all aspects of disease activity to be measured.

The initial phase in the development of an index is usually through consensus approach amongst experts in the field to determine the items that should be included in the index. This approach was used in the development of the BILAG-2004 index as changes to the index were made through a consensus building process involving members of BILAG. This was supplemented by distributing the index to members of the SLICC (System Lupus International Collaborating Clinics) group and British Society for Rheumatology (BSR) SLE Special Interest Group for their comments and opinions. In total, the opinions of 37 lupus experts (non-BILAG members) from within and outside United Kingdom were obtained. All their comments were reviewed and appropriate changes to the index were made accordingly.

However, face and content validity is not a single study but is continually assessed throughout the validation process, whereby changes to index are made based on the data collected (this is described in detail in the relevant chapters of this thesis).

1.8.2 Reliability

Reliability or repeatability is the ability of the index to reproduce similar results when used repeatedly. Variability in measurements can arise from 3 sources, namely patients, the measurement instrument and raters performing the measurement. Patients vary between each other, and also within themselves, at different points in time. It is this variability that an instrument aims to measure. Variation arising from the measurement instrument (measurement error) is a form of systematic error (bias). Variation may also be the result of the rater using the measurement instrument. This form of variability may be within the same rater (intra-rater variability) or between raters (inter-rater variability).

Inter-rater variability occurs when two or more raters assess the same patient within a very short period of time during which the status of the patient's condition has not changed. Intra-rater variability occurs when a single rater assesses a patient with a stable condition at different time points. This is usually difficult to assess due to knowledge bias as the rater would remember the previous assessment if the time interval between assessments is short. However, if the interval between assessments is considerably longer to avoid this bias, it is very likely that the patient's condition will have changed. Intra-rater variability will contribute to inter-rater variability, hence if inter-rater reliability is good then intra-rater reliability can also be assumed to be good as well (Brennan et al. 1992). It is for this reason that only inter-rater reliability will be assessed for the BILAG-2004 index.

The most common statistical method for assessing inter-rater reliability of categorical scale is the kappa statistic (Cohen 1960). The formula for kappa is as follows:

$$\frac{\text{Actual agreement (\%)} - \text{Expected agreement (\%)}}{100 - \text{Expected agreement (\%)}}$$

The numerator is the actual agreement beyond chance while the denominator represents potential agreement beyond chance. Therefore, kappa expresses the actual agreement beyond chance as a proportion of potential agreement beyond chance. Kappa provides an overall measure of agreement between raters and has a value ranging from -1 to 1. There is no absolute interpretation of the kappa values but some guidelines have been suggested (Table 1-3) (Landis et al. 1977). Generally, a value of at least 0.40 is considered to represent acceptable agreement.

Table 1-3: Interpretation of kappa statistic.

Kappa value	Agreement
< 0.0	Poor
0.00 – 0.20	Slight
0.21 – 0.40	Fair
0.41 – 0.60	Moderate
0.61 – 0.80	Substantial
0.81 – 1.00	Almost Perfect

However, kappa is only concerned with agreement and does not take into account partial agreement or different degrees of disagreement. All degrees of disagreement are treated the same, as complete disagreement. A typical table of agreement is shown in Table 1-4. The expected frequency in a cell is calculated as the product of the total in the relevant column and the total in the relevant row, divided by the grand total. Using the table above as an example, $E_{11} = R_1 \times C_1/N$, $E_{12} = R_1 \times C_2/N$, $E_{21} = R_2 \times C_1/N$ and so on. The diagonal cells of the table (E_{11}, E_{22}, E_{33} and E_{44}) represent perfect agreement and the further away from this diagonal, the greater is the degree of disagreement. As the BILAG-2004 index is an ordinal

scale index with decreasing level of activity from Grade A to Grade E, it would be preferable to take into account different levels of disagreement. As an example, the disagreement between Grade A and Grade B is far less severe than the disagreement between Grade A and Grade D. For this purpose, a weighted kappa statistic is used to adjust for the different levels of disagreement (Cohen 1968). The expected frequency for each cell is multiplied by a weight and the weight ranges from 0 to 1. A weight of 1 represents perfect agreement while 0 represents total disagreement and any values in between represent partial agreement with larger values indicating lesser degree of disagreement.

Table 1-4: Typical table of agreement.

		Rater 2				Total
		A	B	C	D/E	
Rater 1	A	E_{11}	E_{12}	E_{13}	E_{14}	R_1
	B	E_{21}	E_{22}	E_{23}	E_{24}	R_2
	C	E_{31}	E_{32}	E_{33}	E_{34}	R_3
	D/E	E_{41}	E_{42}	E_{43}	E_{44}	R_4
Total		C_1	C_2	C_3	C_4	N

Another method for assessing reliability is with intra-class correlation coefficient (ICC) (Shrout et al. 1979). This is derived from analysis of variance and is the ratio of variance between patients to the total variance. ICC is very similar to kappa statistics as ICC and weighted kappa (using quadratic weights) are statistically equivalent (Fleiss et al. 1973). Although this is primarily designed for continuous data, it can be use for ordinal data by transforming the data into a continuous data. The initial reliability study of the preliminary BILAG-2004 index had used this statistical method to assess reliability.

1.8.3 Construct and Criterion Validity

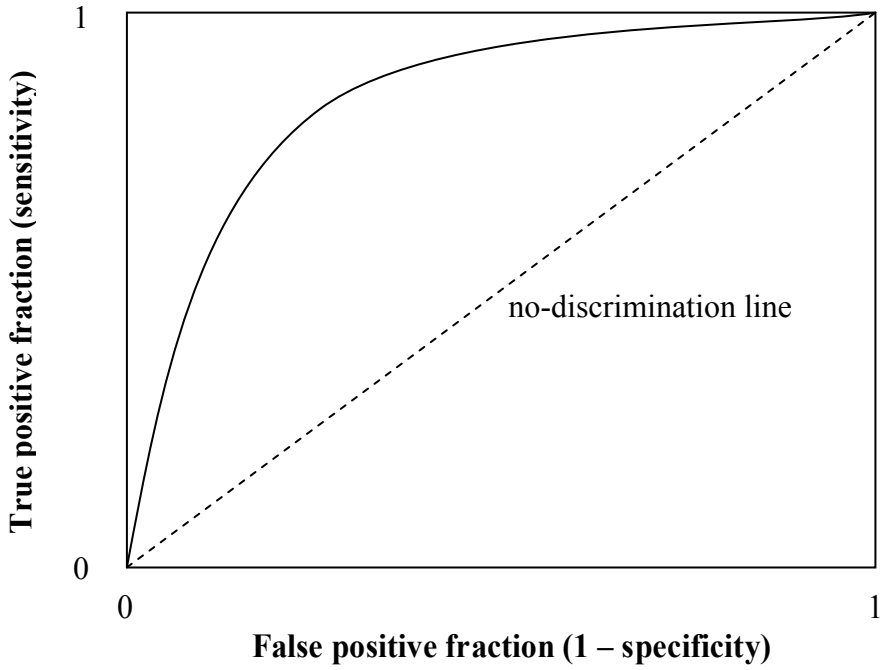
Criterion validity determines if this index agrees with the gold standard. As there is no gold standard for SLE disease activity, change in therapy is used as the reference standard (also known as the criterion) for the validation of the BILAG-2004 index. Construct validity determines if this index correlates well with other surrogate measures of disease activity, known as constructs. These constructs are not the gold standard, but they are correlates of disease activity. A cross sectional study is required for the assessment of construct and criterion validity.

For construct validity, statistical analysis is used to demonstrate the correlation between the index and the surrogate measures (constructs). On the other hand, it is the agreement between the index and the criterion (reference) that is being assessed in criterion validity, by determining the sensitivity, specificity, positive predictive value and negative predictive value of the index. Sensitivity is the percentage of patients with active disease that are correctly classified by the index (or true positive rate) while specificity is the percentage of patients without active disease correctly classified by the index (or true negative rate). Positive predictive value is the probability of a patient, classified as having active disease by the index, really does have active disease while negative predictive value is the probability of a patient, classified as not having active disease by the index, really does have inactive disease.

Receiver operating characteristic (ROC) curve is also commonly used to assess the performance of an index in the analysis for criterion validity (Zweig et al. 1993). The ROC curve is a plot of true positive fraction (sensitivity) against false positive fraction (1 – specificity) (Figure 1-3). An index with perfect discrimination between active and inactive disease states has an ROC plot that passes through the upper left corner where the true

positive fraction is 1.0 and the false positive fraction is 0. The no-discrimination line is a 45 degrees diagonal line from the lower left corner to the upper right corner and this represents the plot for an index that cannot discriminate between the two states. The closer the plot is to the upper left corner, the better is the performance of the index in discriminating between active disease and inactive disease. Another way to quantify the performance of the index is to calculate the area under the ROC curve which provides a quantitative expression of how close the curve is to the perfect one (area = 1.0).

Figure 1-3: Typical receiver operating characteristic plot.



1.8.4 Sensitivity to Change

Sensitivity to change or responsiveness of an index implies the ability of the index to change with time. There are two forms of responsiveness: internal responsiveness and external responsiveness.

Internal responsiveness is defined as the ability of an index to change over a particular time frame. Traditional statistical methods that are used to assess sensitivity to change of

several SLE disease activity indices fall into this category and they include paired t-test, effect size, standardised response mean and Guyatt's responsiveness index. The main disadvantage of these methods is that the changes in the index do not relate to changes on an external measure at individual patient level. Apart from Guyatt's responsiveness index, the statistics only examine the extent of change in the index between two time points. Hence, statistically significant change may occur without corresponding change in clinical status and these changes may not be clinically meaningful or relevant. Moreover, the comparison is made at the population level and may not reflect clinical change at the individual patient level. Furthermore, comparison across different studies is difficult as the statistics used are not independent of study design (Husted et al. 2000).

External responsiveness is the extent to which changes in the index over time relate to corresponding changes in an external reference measure. It is essential that the selected external reference measure represent an accepted indicator of change in the patient's status. Therefore, it is the clinically meaningful change that is being studied. As the statistics are based on correlational approach, they characterise the relationship between change in the index and change in the external reference at the individual patient level. Therefore the result is generalisable across studies, that is the same relationship should be observed in another study of similar patients, allowing for comparison. The common statistical approaches used are receiver operating characteristic method, regression models and correlation analyses (Husted et al. 2000).

It is important that the change in the index over time reflects actual change in the clinical status of the patient when the index is used to measure transition in clinical states, and this affects the number of patients that are required to detect significant differences between

treatment arms in a clinical trial. Therefore, the more robust external responsiveness method is the preferred method to assess sensitivity to change of an index.

1.8.5 Predictive Validity

This is the ability of the index to predict the occurrence of a future outcome. This entails a prospective longitudinal study and the outcome of interest is damage as measured by SLICC/ACR damage index. This longitudinal study is designed to determine if greater cumulative exposure to active disease according to the BILAG-2004 index would lead to development of damage in an inception cohort of SLE patients.

1.8.6 Rasch Analysis

Measurement scales are commonly used in medical sciences and most of them are ordinal scales as they are made up of qualitative items. Traditionally, the rating scale responses of these items are assigned discrete numbers of increasing value to reflect the increasing quantity of the qualitative responses. Subsequently, the scores for all the items are summated into a total or raw score. This raw score is subsequently used as the 'measure' in statistical analyses. Two assumptions are made when this is done. Firstly, each item is assumed to contribute equally to the measure. This implies that all the items are of equal importance in the assessment of the construct. Secondly, all the items are measured on the same linear scale and the distance between each rating scale point is uniform both within and across items. Unfortunately, the validity of these two assumptions has not been tested using traditional approaches. The Rasch analysis provides the only means to assess these assumptions.

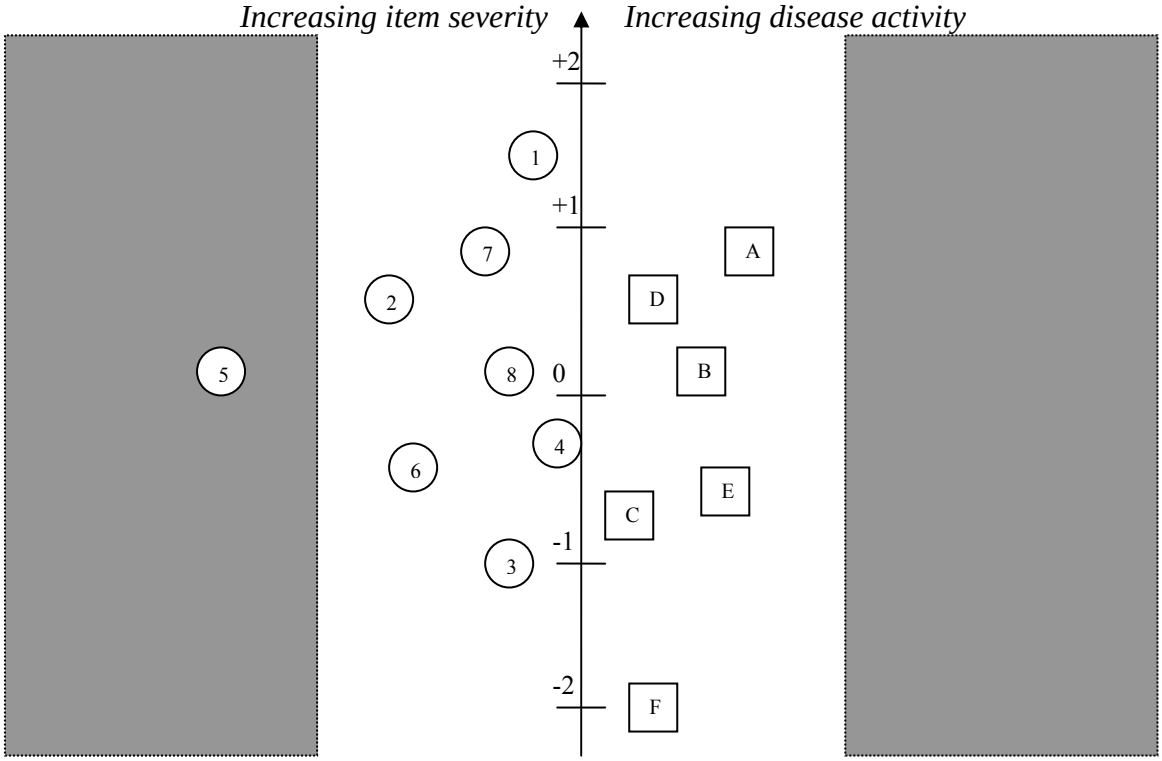
The Rasch model is a probabilistic mathematical model that is used to assess measurement scale scientifically (Andrich 1988; Rasch 1980). It assumes that the probability of a particular response in an item is a logistic function of the severity of the item and the patient's disease activity. Its most basic model is the dichotomous model which is represented by the following equation:

$$\text{logit} = \ln\left(\frac{P}{1-P}\right) = \beta - \delta$$

where P is the probability of a patient's given response to an item, β is the patient's disease activity and δ is the severity of the item. It is the logistic transformation of the probabilities that gives rise to a linear scale, with logit being the basic unit of this scale. An implicit assumption of this model is that there is a hierarchical ordering of the items in the index, which means that the data forms an ordinal scale. This model has the properties of fundamental measurement which are additivity of the scale (logit being the basic unit), and specific objectivity as the equation allows for item severity and patient's disease activity to be independent of each other (item-person separation). The concept of Rasch measurement can be represented by a map of items (within a scale) and patients as shown in Figure 1-4. Using the Rasch model, the item and person data undergo logarithmnic transformation which will convert ordinal data into a linear (logit) scale. This transformation will provide the estimates for item severity and patient's disease activity in logit scale, which allow both items and patients to be displayed on the same map (according to its estimated value). The map in Figure 1-4 shows the relative location of 6 patients (in squares labelled A to F) with differing levels of disease activity and 8 items (in circles labelled 1 to 8) with differing severity. The more severe items and patients with more active disease are located further towards the top of the map (more positive logit value). Therefore, item 1 has the highest severity (most difficult to achieve) while item 3 being at the other end of the map is the least severe (easiest to

achieve). Items 5 and 8 have similar level of severity as they are on the same location on the logit scale. Similarly, patient A has the most active disease (being at the top of the map) followed by D, B, E, C and F. Data that are performing to the model expectation would be located in the white zone and data that do not fit the model (as with item 5) will be in the shaded area.

Figure 1-4: Representation of Rasch measurement using item-person map. (square represents person, circle represent item, shaded areas represent data that do not fit the Rasch model)



Apart from that, this model also confers unidimensionality whereby the index measures only one attribute at a time. Measurement scales should strive to achieve this property as multidimensional scales have a number of attributes merged into a single generic score that makes interpretation difficult. This model is also used to assess internal construct validity of the index as unidimensionality infers that the items in the index are measuring the same attribute.

Unlike traditional statistical analysis where the statistical model tries to fit the data, the reverse is the case with Rasch analysis as the data needs to conform to the model. The original Rasch model was designed for dichotomous data, but since its publication, the model has been expanded for use in other situations.

1.9 Objectives of Research

The objectives of my thesis were to develop and validate the BILAG-2004 index for the assessment of disease activity in SLE with the following aims:

(a) to ensure that this index has face and content validity

(b) to determine the inter-rater reliability of this index in routine clinical practice

(c) to determine the construct and criterion validity of this index

(d) to determine if this index is sensitive to change

(e) to determine the predictive validity of this index

(f) to determine if this index fits the Rasch model

Chapter 2 General Methodology

The validation of the BILAG-2004 index has involved the following studies:

1. Reliability study in routine clinical practice
2. Cross sectional study (construct and criterion validity)
3. Sensitivity to change study
4. Predictive validity study
5. Rasch analysis

All these studies involved multiple centres in United Kingdom and had received multi-centre research ethical approval from Hull and East Riding Research Ethics Committee as well as approval from local research ethics committees of all participating centres. These studies were carried out in accordance with the Helsinki Declaration and written consent was obtained from patients. Patients with SLE who satisfied the revised ACR criteria for classification of SLE were recruited from both in-patients and out-patients (Hochberg 1997; Tan et al. 1982). Patients were excluded from the study if they were pregnant, under age of 18 years or unable to give valid consent.

The specific details of each study are described in subsequent relevant chapters. This chapter provides descriptions of certain aspects of methodology and analysis which are common to multiple chapters.

2.1 Definition of Change in Therapy

Change in therapy was used as the reference standard for disease activity in the cross sectional study (for criterion validity) and in the sensitivity to change study (which is a longitudinal study). The definition used for both studies was similar, however there was a slight but important difference. In the cross sectional study, change in therapy was the difference in treatment after the patient was assessed as compared to the treatment that the patient was on prior to the assessment. In other words, it was the change in the treatment following the assessment. In contrast, change in therapy in the sensitivity to change study was the difference in treatment after the patient was assessed at the index visit, as compared to the therapy after the previous visit (or change in treatment between two consecutive visits).

The medications of interest were immunosuppressives, antimalarials, corticosteroids, biological therapy, topical corticosteroids, topical immunosuppressives, intravenous immunoglobulins, plasmapheresis, anticoagulation, prasterone, dapsone, thalidomide and retinoids. NSAIDs were not included, as they were commonly used for several other indications (for pain relief in non-SLE conditions) and could be obtained over the counter as non-prescription medication.

Three categories of change were defined, namely ‘no change’, ‘increase in therapy’ or ‘decrease in therapy’. Change in therapy was classified into two categories for the purpose of analysis, namely ‘increase in therapy’ and ‘no increase in therapy’. Therefore, no increase in therapy represents a combination of ‘no change’ and ‘decrease in therapy’. This was chosen as increase in therapy was seen as a marker of disease activity. A robust definition for change in therapy was used.

Increase in therapy was defined as any increase in the medications of interest regardless of any concomitant reduction in other medications. Decrease in therapy was

defined as any decrease in the medications of interest without any concomitant increase in other medications. However, change in therapy was not just a simple change in the dose of the medications. The following special circumstances had to be taken into account:

1. Dosing levels based on body weight
2. Step-down switch of immunosuppressive therapy
3. Gradual escalation of immunosuppressive therapy following initiation
4. Increase in immunosuppressive therapy for steroid-sparing effect
5. Initiation of anticoagulation for active SLE disease
6. Reduction or discontinuation of therapy due to side effects

Dosing Levels

For some immunosuppressives, different dosing levels based on body weight were considered in the definition of change in therapy (Table 2-1). A change in therapy was deemed to have occurred when there had been a change in the dosing level of these medications. These levels were based on clinical judgement and experience.

Table 2-1: Dosing levels of medications used for defining change in therapy.

Medications	**Level 1**	**Level 2**	**Level 3**
Azathioprine	< 1 mg/kg/d	1 - 2.4 mg/kg/d	≥ 2.5 mg/kg/d
Mycophenolate	< 2 g/d	2 g/d	> 2 g/d
Ciclosporin A	< 2 mg/kg/d	2 - 3 mg/kg/d	> 3 mg/kg/d
Tacrolimus	< 0.10 mg/kg/d	0.10 - 0.15 mg/kg/d	> 0.15 mg/kg/d
Methotrexate	< 10 mg/wk	10 - 15 mg/wk	> 15 mg/wk
Oral Cyclophosphamide	< 1 mg/kg/d	1 - 2 mg/kg/d	> 2 mg/kg/d

For medications that were not listed in Table 2-1, a simple change in dose would constitute a change in therapy.

Change in Immunosuppressives

A switch in immunosuppressive therapy was generally considered as an increase in therapy except in the situation of changing cyclophosphamide to azathioprine, methotrexate or ciclosporin. This was because it was common practice to make such a change once the disease was under control, as prolonged cyclophosphamide therapy was associated with significant toxicity (step-down phase). In fact, this step-down phase was equivalent to a reduction in therapy as the discontinuation of cyclophosphamide was considered as a decrease in therapy while the initiation of the other immunosuppressive was not considered an increase in this circumstance. In the case of change from cyclophosphamide to mycophenolate mofetil, the situation was clarified with the local investigator as to whether the change was a result of failed cyclophosphamide (indicating increase in therapy) or as a step-down phase (indicating no increase in therapy).

Immunosuppressives for Steroid-Sparing Effect

If an immunosuppressive agent was started for its steroid-sparing effect, this was not considered to be an increase in therapy.

Anticoagulation for Active Disease

Anticoagulation needed to be initiated for reason of active disease (which was clarified with the local investigator) and in the presence of immunosuppressives or high dose corticosteroids before it was considered an increase in therapy.

Escalation of Immunosuppressive Dose

As most immunosuppressives have potential toxicity, it is common practice to start at a low dose and gradually escalate to the target dose. To take this into account, any increase in the dose of immunosuppressives within the first 3 months of initiation was considered as part of an escalation plan to achieve the target dose and not as an increase in therapy. Similarly, it was also common practice to reduce corticosteroids dose gradually during this period as part of the escalation plan. Therefore, any concomitant reduction in corticosteroids dose during the escalation phase was not considered as a reduction in therapy.

Reduction of Therapy due to Side Effects

If any medication was decreased or discontinued due to side effects, this was not considered to be a reduction in treatment.

2.2 Common Statistical Analysis

Overall BILAG-2004 and overall Classic BILAG scores were used in the analysis in several of the studies. These overall scores were determined by the highest score achieved by any system in the respective index. BILAG-2004 and Classic BILAG scores of D and E were combined together as both of them indicate inactivity. Therefore, four categorical overall scores were possible (A, B, C and D).

The majority of the patients in the cross sectional and sensitivity to change studies contributed more than one observation to the respective studies as there were repeated observations over time from the same patients. In both of these studies, patients who had more active and severe disease were more likely to be seen and assessed more frequently, thereby contributing more observations compared to patients with stable or mild disease. Therefore,

observations from the same patient were more likely to be correlated and independence of these observations could not be assumed. To account for the repeat observations from the same patients, robust variance estimation (Huber/White/sandwich variance estimator) was used instead of the standard variance estimation in the logistic regression analyses (Williams 2000).

All statistical analyses were performed using Stata for Windows (Stata Corporation, Texas). The specific details of the analysis used in the various validation studies will be described in subsequent relevant chapters.

Chapter 3 Reliability Study in Routine Clinical Practice

3.1 Abstract

Objectives:

1. To determine the inter-rater reliability of the BILAG-2004 and SLEDAI-2000 indices
2. To compare the ability of the BILAG-2004 and SLEDAI-2000 indices to capture disease activity

Methods:

SLE patients were recruited from 11 centres and were assessed by 2 raters separately in routine clinical practice. Disease activity was determined using the BILAG-2004 index. Two exercises were performed: changes were made after the first exercise (E1) and training was provided to the raters before the second exercise (E2). SLEDAI-2000 index was also used in E2. E1 and E2 involved 12 and 14 raters respectively. Reliability was assessed using level of agreement, kappa statistic, intraclass correlation coefficient (ICC) and analysis of disagreement. Ability to detect disease activity was assessed by determining the number of patients with high activity on the BILAG-2004 index (overall score A or B) but low SLEDAI-2000 score (< 6) and number of patients with low activity on the BILAG-2004 index (overall score C, D or E) but high SLEDAI-2000 score (≥ 6). Treatment of these patients were analysed and increase in therapy was used as the reference standard for active disease.

Results:

Ninety seven patients were recruited for both exercises. There was improvement in agreement, kappa statistics and ICC, with reduction in major disagreements from E1 to E2. Reliability was further improved by removing poorly performing items. Only 93 patients in

E2 had SLEDAI-2000 assessments. There were 43 patients (46.2%) with a difference in SLEDAI-2000 score between the 2 raters and this difference was ≥ 4 in 19 patients (20.4%). Agreement for each of the items in SLEDAI-2000 was between 81.7 – 100%. Thirty five patients (37.6%) had high activity on BILAG-2004 but low SLEDAI-2000 score, of which 48.6% had their therapy increased. There were 5 patients (5.4%) with low activity on BILAG-2004 but high SLEDAI-2000 score.

Conclusions:

The BILAG-2004 and SLEDAI-2000 indices are reliable to assess SLE activity. SLEDAI-2000 index appears to be less able to detect active disease requiring increased therapy than the BILAG-2004 index.

3.2 Objectives

1. To determine the inter-rater reliability of the BILAG-2004 index in routine clinical practice
2. To determine the inter-rater reliability of the SLEDAI-2000 index in routine clinical practice
3. To compare the ability of the BILAG-2004 and SLEDAI-2000 indices to capture disease activity

3.3 Introduction

As discussed in Chapter 1, SLE is a complex multi-system autoimmune disease with diverse immunological and clinical manifestations. Assessment of disease activity poses a challenging problem as any organ system can be affected and it is well known that SLE may mimic manifestations of other diseases. As there is no single biomarker that adequately reflects disease activity, numerous composite clinical indices have been developed for the assessment of disease activity (Liang et al. 1988). The two most commonly used disease activity indices in clinical studies are the Classic BILAG and SLEDAI indices. Over time, several deficiencies were noted with both of these indices and they were revised resulting in the BILAG-2004 and SLEDAI-2000 indices, respectively.

To-date, almost all the validation studies of the SLEDAI index involved the original index and SLEDAI-2000 had not been fully validated (Chang et al. 2002a; Chang et al. 2002b; Ward et al. 2000; Gladman et al. 1994; Guzman et al. 1992; Petri et al. 1992; Gladman et al. 1992; Liang et al. 1989). The SLEDAI-2000 index was only validated retrospectively

against the SLEDAI index which showed that SLEDAI-2000 correlated with its predecessor (Gladman et al. 2002).

The preliminary BILAG-2004 index had been subjected to an initial validation with reliability study (Isenberg et al. 2005). Two reliability exercises were performed, in which 8 patients were assessed on each occasion by 8 rheumatologists in a specially designed research clinic. The results showed that most of the systems had good reliability but there was concern regarding the performance of the Musculoskeletal system, which was most likely related to inadequacies in the glossary. Changes were made to improve the glossary after the study.

In this chapter, I will present the results of a much larger inter-rater reliability study of the BILAG-2004 and SLEDAI-2000 indices in routine clinical practice. The performance of both of these indices was also compared with regards to their ability to detect disease activity requiring an increase in therapy.

3.4 Methods

This was a multi-centre cross sectional study involving 11 centres in the United Kingdom. Recruited SLE patients were assessed separately by the local rheumatologist and an external physician (CSY). This study was performed in the setting of routine clinical practice and the medical records were available to both of the raters. Disease activity was assessed using the BILAG-2004 index. Raters were also asked to record any anomalies or difficulties encountered with the index during the study. After the study in each centre, there was discussion between the raters on the difference in scoring between the two raters, and issues related to the face and content validity of the index and its glossary.

Two reliability exercises were performed. During the first exercise (E1), several issues with the glossary of the index were identified and it was noted that several raters were

insufficiently familiar with the index resulting in errors on using the index. Following this, changes were made to the glossary of the BILAG-2004 index. A second exercise (E2) using the revised BILAG-2004 index was planned a few months later and it was decided that for this exercise, SLEDAI-2000 index (Gladman et al. 2002) would also be used to assess disease activity as a comparison. Prior to the second exercise, training was provided to the raters to familiarise them with both the revised BILAG-2004 and SLEDAI-2000 indices. Twelve patients were involved in both of the exercises. The first exercise involved 12 raters while the second exercise involved 14 raters.

3.5 Analysis

For the purpose of analysis, the BILAG-2004 index scores of D and E were combined together as both of them indicate inactivity. For each system, inter-rater reliability of the BILAG-2004 index was assessed using kappa statistic (unweighted and weighted) based on a simple two way tabulation of external versus local rater scores (Cohen 1968; Cohen 1960). In addition, a pooled within centre intraclass correlation coefficient (ICC) was calculated for each system through the use of numerical scores for each categorical grade (A=9, B=3, C=1 and D/E=0) (Stoll et al. 1997). The ICC used in this analysis was a slight generalisation of the ICC equation [2,1] given by Shrout and Fleiss (Shrout et al. 1979). An overall ICC for the index was calculated using the summated numerical scores of all systems. Calculations of ICC were performed by Dr Elizabeth Allen.

Confidence intervals (CI) were calculated for kappa statistics and ICCs. As BILAG-2004 scoring has more than 2 categories, a bootstrap technique with 1000 replications was used to estimate a 95% confidence interval for kappa statistic. Bootstrap technique is a computer-intensive method used to estimate parameters and confidence intervals for models,

which require fewer assumptions about the distribution of the data than parametric methods (Davison et al. 2007). It infers the distribution of the population from the distribution of the sample data and derives estimates by making repeat random samples of the same size as the original sample from the data itself. However, there is replacement of the sample using the computer in which any observation can be sampled more than once. This replacement is important to avoid random permutation of the original data that would result in the same estimate as the original data. The confidence intervals are estimated from the variability derived from this sampling.

Percentages of assessments on which both raters agreed were calculated, as was a weighted level of agreement. The weighting used to calculate the weighted level of agreement and weighted kappa statistic reflects the judgement of the raters on the severity of the possible disagreements and is as follows:

Table 3-1: Weighting used to calculate weighted agreement and weighted kappa statistic.

Arbitrary weight		Rater 1			
		A	B	C	D/E
Rater 2	A	1.0	0.75	0.25	0
	B	0.75	1.0	0.5	0
	C	0.25	0.5	1.0	0.5
	D/E	0	0	0.5	1.0

As Grades A and B represented active disease and Grades D and E indicated inactivity, it was decided that this difference in scoring between raters would be considered total disagreement (weighting score of zero). Similarly, with Grade C indicating mild disease, the score difference between raters of A and C was considered as minor agreement of about 25% and hence the weighting of 0.25. This weighting system is similar to the one used in the reliability study of the Classic BILAG index (Hay et al. 1993). An acceptable level of agreement was set

at the level of at least 70%. As for kappa statistic, the acceptable value was that of at least 0.40 (Landis et al. 1977). With ICC and weighted kappa statistic being equivalent statistically, the acceptable value for ICC was the same as kappa statistic (Fleiss et al. 1973).

With active disease being represented by scores of A and B and the other grades (C, D and E) indicating minimal activity, disagreements in scoring between the raters were classified into major disagreement and minor disagreement. Major disagreement in scoring was defined as scores of A by one rater and C or D or E by the other rater, or scores of B by one rater and D or E by the other. Minor disagreement was defined as a score difference of A and B, or score difference of B and C, between raters. Score difference of C and D or E was generally regarded as acceptable level of disagreement.

The reliability of SLEDAI-2000 index was assessed using the level of agreement for each item in the index. The ability to detect disease activity was assessed by determining the number of patients with discordant BILAG-2004 and SLEDAI-2000 scores. These discordant scores can be divided into those with high activity on the BILAG-2004 index (overall score of A or B) but low SLEDAI-2000 score (less than 6), and those with low activity on the BILAG-2004 index (overall score of C or D or E) but high SLEDAI-2000 score (6 or more). The overall BILAG-2004 score for a patient was determined by the highest score achieved by any system in the index. Treatment of these patients was analysed and increase in corticosteroids, antimalarials or cytotoxic therapy was used as the reference standard for active disease. Assessment by the external rater was used for this analysis to avoid rater effect and to minimise bias as treatment decisions were made by the local rater. Analysis with different SLEDAI-2000 cut-off scores for definition of active disease was also performed.

3.6 Results

Ninety seven patients were recruited for both of the exercises. The demographics of the patients are summarised in Table 3-2.

Table 3-2: Demographics of patients recruited into the reliability exercises.

Patient Characteristics	Exercise 1	Exercise 2
Female sex (%)	89.7	90.7
Mean age in years (range)	42.3 (18.5 – 82.2)	43.7 (17.7 – 75)
Ethnicity (%)		
Caucasian	74.2	68.0
Afro-Caribbean	8.2	15.5
South Asian	13.4	11.3
Oriental	0	1.0
Others	4.1	4.1
Mean disease duration in years (range)	9.4 (0 – 32.1)	10 (0 – 34.8)

The distribution of disease activity across the systems for both exercises as assessed by the local physician using the BILAG-2004 index is summarised in Table 3-3 and Table 3-4.

Table 3-3: Distribution of disease activity across systems for Exercise 1 using the BILAG-2004 index as assessed by local physician.

System	A (%)	B (%)	C (%)	D (%)
Constitutional	1.0	11.3	52.6	35.1
Mucocutaneous	5.1	19.6	25.8	49.5
Neuropsychiatric	8.2	1.0	24.7	66.0
Musculoskeletal	4.1	13.4	42.3	40.2
Cardiorespiratory	3.1	8.2	1.0	87.6
Gastrointestinal	0	1.0	1.0	97.9
Ophthalmic	0	2.1	2.1	95.9
Renal	1.1	7.7	14.4	78.4
Haematological	0	3.4	47.4	46.4

Table 3-4: Distribution of disease activity across systems for Exercise 2 using the BILAG-2004 index as assessed by local physician.

System	A (%)	B (%)	C (%)	D (%)
Constitutional	1.0	9.3	51.5	38.1
Mucocutaneous	3.1	23.7	34.0	39.2
Neuropsychiatric	4.1	8.2	16.5	71.1
Musculoskeletal	4.1	18.6	38.1	39.2
Cardiorespiratory	0	9.3	3.1	87.6
Gastrointestinal	0	1.0	0	99.0
Ophthalmic	0	1.0	1.0	97.9
Renal	4.1	4.1	5.2	86.6
Haematological	0	10.3	24.7	64.9

The majority of the patients in both exercises had relatively low disease activity (Grades C or D) and this was particularly evident in the Gastrointestinal and Ophthalmic systems. This remained an issue despite targeted recruitment of patients with active disease for the second exercise in view of the clustering of low activity in the first exercise. Even in patients with evidence of active disease, most of them had only one system involved while the other systems were usually unaffected or minimally affected.

3.6.1 Inter-rater Agreement and Reliability of the BILAG-2004 Index

The results of the first exercise are summarised in Tables 3-5 to 3-7. The details are shown in Appendix 5. The overall ICC for the BILAG-2004 index was 0.45 (95% CI: 0.31, 0.58). There was poor level of agreement (< 70%) in the Constitutional, Mucocutaneous and Musculoskeletal systems, while the kappa was low (< 0.40) in Constitutional, Neuropsychiatric, Musculoskeletal, Gastrointestinal and Ophthalmic systems, and, the estimated ICC was low (< 0.40) in Constitutional, Neuropsychiatric, Cardiovascular, Gastrointestinal and Ophthalmic systems. The low kappa and ICC values despite excellent

agreement in Ophthalmic and Gastrointestinal systems were due to clustering of inactivity in these systems in most patients.

The presence of large number of major disagreements (0.44 per patient) was rather worrying with the majority of them in Constitutional, Mucocutaneous, Neuropsychiatric, Musculoskeletal and Cardiorespiratory (Table 3-7). Further analysis into the disagreements revealed that most of them (98.6%) could be attributed to non-scale factors (not related to the index). Almost all the non-scale factors were related to the raters. It was evident that there were numerous recording errors (misclassification), such as recording of features that were not due to SLE (for example painful knee effusion from osteoarthritis), and errors in interpretation of the terms used in the index (usually the result of not referring to the glossary), for example scoring moderate inflammatory arthritis as severe inflammatory arthritis. These errors could be attributed to insufficient familiarity with the BILAG-2004 index.

Table 3-5: Agreement for each system in the BILAG-2004 index for Exercise 1.

System	Agreement %	Weighted Agreement %
Constitutional	48.5	69.6
Mucocutaneous	69.1	80.4
Neuropsychiatric	74.2	84.0
Musculoskeletal	51.6	72.4
Cardiorespiratory	90.7	92.0
Gastrointestinal	96.9	97.4
Ophthalmic	95.9	96.9
Renal	96.9	98.5
Haematological	91.5	95.7

Table 3-6: Kappa statistic (unweighted and weighted) and ICC of each system in the BILAG-2004 index for Exercise 1.

System	Kappa (95% CI)	Weighted Kappa (95% CI)	ICC (95% CI)
Constitutional	0.18 (0.05, 0.31)	0.18 (0.06, 0.33)	0.09 (0, 0.21)
Mucocutaneous	0.50 (0.36, 0.65)	0.55 (0.41, 0.71)	0.40 (0.26, 0.53)
Neuropsychiatric	0.36 (0.18, 0.55)	0.38 (0.20, 0.59)	0.34 (0.20, 0.47)
Musculoskeletal	0.26 (0.13, 0.39)	0.35 (0.22, 0.49)	0.43 (0.30, 0.56)
Cardiorespiratory	0.53 (0.25, 0.78)	0.58 (0.33, 0.82))	0.18 (0.06, 0.32)
Gastrointestinal	-0.01 (-0.02, 0)	-0.01 (-0.03, 0)	0 (0, 0.11)
Ophthalmic	0 (0, 0)	0 (0, 0)	0 (0, 0.11)
Renal	0.91 (0.79, 1)	0.94 (0.84, 1)	0.97 (0.96, 0.98)
Haematological	0.84 (0.71, 0.94)	0.85 (0.74, 0.93)	0.75 (0.65, 0.82)

Table 3-7: Analysis of disagreement in scoring of the BILAG-2004 index between raters in Exercise 1.

Disagreements	Total	Systems
A and C or D	13	Neuropsychiatric 5 Mucocutaneous 4 Cardiorespiratory 3 Constitutional 1
B and D	30	Constitutional 8 Musculoskeletal 8 Mucocutaneous 6 Cardiorespiratory 3 Gastrointestinal 2 Ophthalmic 2 Neuropsychiatric 1
A and B	6	Musculoskeletal 4 Mucocutaneous 1 Cardiorespiratory 1
B and C	22	Musculoskeletal 11 Mucocutaneous 6 Constitutional 3 Cardiorespiratory 1 Haematological 1

Several issues with the glossary and definitions were identified and changes were recommended as follows:

1. There was no clear guidance in the form and glossary on the differentiation of activity and damage due to SLE. This resulted in misclassification of items, with features that were due to damage being recorded as disease activity. Hence, a clear indication that only features due to activity are to be recorded was needed in the form and glossary. Apart from that, guidance on the differentiation of activity from damage was provided.
2. New episodes of manifestations that occurred in the last 4 weeks and not in the previous 4 weeks were being recorded as 'new' even if they had improved significantly prior to the assessment. It was noted that this would result in over-scoring of the given manifestations as it would be uncommon for physicians to increase

therapy in this situation. It was therefore more appropriate that in this situation, the item should be recorded as ‘improving’ instead of ‘new’, which would result in a lower score provided that the item fulfilled the criteria for improving.

3. Many items in the index had not been adequately qualified or quantified in the glossary and this had resulted in different thresholds or interpretations being used by individual raters. The definitions of items in the index were re-examined and changes were made accordingly to the glossary.
4. There was confusion in the interpretation of the term polyarthritis which was used in the item ‘severe polyarthritis’. It was commonly misunderstood to mean the conventional definition which was involvement of 5 or more joints, while in the BILAG-2004 index this referred to severe involvement of 2 or more joints. Therefore, the item ‘severe polyarthritis’ was rephrased into ‘severe arthritis’.

In view of the above results in Exercise 1 which was less than satisfactory, a repeat exercise was conducted using the revised BILAG-2004 index incorporating the above changes. Training was provided to the raters to familiarise them with the index before the second exercise. The training was in the form of correspondence and a short briefing before the start of the second exercise in each centre, to ensure familiarity with the items and the glossary definitions.

The results of the second exercise are summarised in Tables 3-8 to 3-10, with details in Appendix 4.

Table 3-8: Agreement for each system in the BILAG-2004 index for Exercise 2.

Systems	Agreement %	Weighted Agreement %
Constitutional	53.6	75.3
Mucocutaneous	72.2	84.3
Neuropsychiatric	80.4	86.9
Musculoskeletal	56.7	77.3
Cardiorespiratory	90.7	92.3
Gastrointestinal	99.0	99.0
Ophthalmic	99.0	99.0
Renal	95.9	97.9
Haematological	88.7	94.3

Table 3-9: Kappa statistic (unweighted and weighted) and ICC of each system in the BILAG-2004 index for Exercise 2.

Systems	Kappa (95% CI)	Weighted Kappa (95% CI)	ICC (95% CI)
Constitutional	0.24 (0.12, 0.39)	0.31 (0.18, 0.44)	0.63 (0.51, 0.74)
Mucocutaneous	0.58 (0.46, 0.70)	0.65 (0.52, 0.76)	0.46 (0.31, 0.60)
Neuropsychiatric	0.46 (0.26, 0.65)	0.50 (0.30, 0.70)	0.80 (0.71, 0.86)
Musculoskeletal	0.34 (0.20, 0.48)	0.47 (0.32, 0.59)	0.55 (0.41, 0.66)
Cardiorespiratory	0.44 (0.12, 0.76)	0.45 (0.08, 0.75)	0.38 (0.23, 0.53)
Gastrointestinal	0 (0, 0)	0 (0, 0)	0.08 (0, 0.23)
Ophthalmic	0.66 (0, 1)	0.49 (0.33, 0.82)	0.20 (0.05, 0.35)
Renal	0.80 (0.57, 0.96)	0.88 (0.71, 0.97)	0.99 (0.98, 0.99)
Haematological	0.79 (0.66, 0.89)	0.82 (0.71, 0.91)	0.86 (0.80, 0.90)

Table 3-10: Analysis of disagreement in scoring of the BILAG-2004 index between raters in Exercise 2.

Disagreements	Total	Systems
A and C or D	1	Musculoskeletal 1
B and D	26	Constitutional 3 Musculoskeletal 3 Mucocutaneous 5 Cardiorespiratory 6 Gastrointestinal 1 Ophthalmic 1 Neuropsychiatric 7
A and B	7	Musculoskeletal 3 Mucocutaneous 3 Neuropsychiatric 1
B and C	32	Musculoskeletal 13 Mucocutaneous 10 Constitutional 7 Haematological 2

The overall ICC for this second exercise was 0.67 (95% confidence interval: 0.54, 0.76). There was low kappa in the Constitutional, Gastrointestinal and Ophthalmic systems while the level of agreement for all systems was acceptable when weighted agreement was considered. The ICC was low in the Cardiorespiratory, Gastrointestinal and Ophthalmic systems. The low kappa and/or ICC values despite good agreement in Ophthalmic and Gastrointestinal systems were due to clustering of inactivity in these systems in most patients. In general, there was improvement in the level of agreement, kappa and ICC from Exercise 1 to Exercise 2. There were also far fewer major disagreements (0.28 per patient) than in Exercise 1.

It was also noted in both exercises that some items in the index consistently had rather poor agreement between raters, and these items were:

1. Fatigue in the Constitutional system

2. Lupus migraine/cluster/tension headache in the Neuropsychiatric system
3. Mood disorder (depression/mania) in the Neuropsychiatric system
4. Anxiety disorder in the Neuropsychiatric system
5. Arthralgia/Myalgia in the Musculoskeletal system

A repeat analysis of these three systems (Constitutional, Neuropsychiatric and Musculoskeletal) with the above items removed was performed. The results are summarised in Table 3-11 and Table 3-12

With the removal of these items, there was substantial improvement in the level of agreement in both of the exercises. The kappa statistic and ICC values generally improved slightly for Exercise 1. As for Exercise 2, there were large increases for both kappa statistic and ICC values for the Neuropsychiatric system, but these modifications resulted in a large decrease in the ICC value of the Constitutional system due to increased homogeneity of the population studied with removal of the item fatigue. There were noticeable increases in the width of some confidence intervals after the modifications which changed the extent of the heterogeneity in the observed scores. More importantly, there were far fewer major disagreements in the Constitutional and Neuropsychiatric systems. However, in the Musculoskeletal system, there was a 2 to 3 fold increase in the number of major disagreements following the removal of the arthralgia/myalgia item. Therefore, the items fatigue, lupus migraine/cluster/tension headache, mood disorders and anxiety disorder were removed from this index while the item arthralgia/myalgia was retained.

Table 3-11: Analysis of level of agreement, kappa statistic, ICC and disagreements between raters in Exercise 1 for the Constitutional, Neuropsychiatric and Musculoskeletal systems before and after removal of some items (fatigue, lupus migraine/cluster/tension headache, mood disorder, anxiety disorder and arthralgia/myalgia)

Systems	Before Modifications	After Modifications
Constitutional		
Agreement (unweighted, %)	48.5	82.5
Kappa (unweighted)	0.18	0.19
Kappa (weighted)	0.18	0.26
ICC (95% CI)	0.09 (0, 0.21)	0.24 (0.11, 0.38)
A and C or D disagreement	1	0
B and D disagreement	8	3
A and B disagreement	0	0
B and C disagreement	3	2
Neuropsychiatric		
Agreement (unweighted, %)	74.2	91.8
Kappa (unweighted)	0.36	0.40
Kappa (weighted)	0.38	0.44
ICC (95% CI)	0.34 (0.20, 0.47)	0.35 (0.21, 0.48)
A and C or D disagreement	5	5
B and D disagreement	1	1
A and B disagreement	0	0
B and C disagreement	0	0
Musculoskeletal		
Agreement (unweighted, %)	51.6	74.2
Kappa (unweighted)	0.26	0.35
Kappa (weighted)	0.35	0.39
ICC (95% CI)	0.43 (0.30, 0.56)	0.41 (0.27, 0.54)
A and C or D disagreement	0	0
B and D disagreement	8	17
A and B disagreement	4	3
B and C disagreement	11	3

Table 3-12: Analysis of level of agreement, kappa statistic, ICC and disagreements between raters in Exercise 2 for the Constitutional, Neuropsychiatric and Musculoskeletal systems before and after removal of some items (fatigue, lupus migraine/cluster/tension headache, mood disorder, anxiety disorder and arthralgia/myalgia)

Systems	Before Modifications	After Modifications
Constitutional		
Agreement (unweighted, %)	53.6	81.4
Kappa (unweighted)	0.24	0.20
Kappa (weighted)	0.31	0.31
ICC (95% CI)	0.63 (0.51, 0.74)	0.37 (0.22, 0.52)
A and C or D disagreement	0	0
B and D disagreement	3	1
A and B disagreement	0	0
B and C disagreement	7	2
Neuropsychiatric		
Agreement (unweighted, %)	80.4	91.8
Kappa (unweighted)	0.46	0.59
Kappa (weighted)	0.50	0.62
ICC (95% CI)	0.80 (0.71, 0.86)	0.60 (0.47, 0.71)
A and C or D disagreement	0	0
B and D disagreement	7	6
A and B disagreement	1	1
B and C disagreement	0	0
Musculoskeletal		
Agreement (unweighted, %)	56.7	77.3
Kappa (unweighted)	0.34	0.42
Kappa (weighted)	0.47	0.49
ICC (95% CI)	0.55 (0.41, 0.66)	0.82 (0.74, 0.87)
A and C or D disagreement	1	1
B and D disagreement	3	14
A and B disagreement	3	3
B and C disagreement	13	2

3.6.2 Inter-rater Agreement of SLEDAI-2000 Index

Only data from 93 patients in Exercise 2 were available for this analysis as there was no SLEDAI-2000 scoring by the local rater in 4 patients. There were 43 patients (46%) with a difference in the total SLEDAI-2000 score between the 2 raters. Of these, 19 patients (20%)

had score difference between raters of 4 or more. There was good level of agreement in the items of SLEDAI-2000, ranging from 81.7% to 100% (Table 3-13). However, all the clinical items with perfect agreement (items 1, 2, 4, 5, 6, 10 and 19) had null score by both raters.

Table 3-13: Level of agreement between raters for items in SLEDAI-2000 index.

Item	Descriptor	Agreement (%)
1	Seizure	100
2	Psychosis	100
3	Organic brain syndrome	98.9
4	Visual disturbance	100
5	Cranial nerve disorder	100
6	Lupus headache	100
7	Cerebrovascular accident	98.9
8	Vasculitis	98.9
9	Arthritis	89.3
10	Myositis	100
11	Urinary casts	100
12	Haematuria	100
13	Proteinuria	100
14	Pyuria	100
15	Rash	84.9
16	Alopecia	81.7
17	Mucosal ulcers	87.1
18	Pleurisy	98.9
19	Pericarditis	100
20	Low complement	100
21	Increased DNA binding	100
22	Fever	98.9
23	Thrombocytopenia	100
24	Leukopenia	100

3.6.3 Ability to Detect Disease Activity

There were 54 patients (58.1%) with high activity according to the BILAG-2004 index (overall score of A or B) and of these, 29 patients (53.7%) had their therapy increased while 5 patients (9.3%) had their therapy reduced. However, there were far fewer patients (24 patients, 25.8%) with high activity according to SLEDAI-2000 index (score of 6 or more) and of these, 14 patients (58.3%) had their therapy increased whereas 2 patients (8.3%) had their therapy reduced.

Thirty five patients (37.6%) had high activity on the BILAG-2004 index but low SLEDAI-2000 score, whereas there were only 5 patients (5.4%) with low activity on the BILAG-2004 index but high SLEDAI-2000 score (Table 3-14). This difference was statistically significant (p=0.015). When data from the local rater were used, there were more patients (41, 44.1%) with high activity on BILAG-2004 but low SLEDAI-2000 score and less patients (4, 4.3%) with low activity on BILAG-2004 but high SLEDAI-2000 score.

Table 3-14: Cross tabulation of BILAG-2004 and SLEDAI-2000 scores for patients recruited into the study.

	SLEDAI-2000 Score ≥ 6	SLEDAI-2000 Score < 6	Total
Overall BILAG-2004 Score of A or B	19	35	54
Overall BILAG-2004 Score of C or D or E	5	34	39
Total	49	44	93

The treatment of these patients with discordant BILAG-2004 and SLEDAI-2000 scores is summarized in Tables 3-15 and 3-16.

Table 3-15: Treatment analysis of patients with high BILAG-2004 and low SLEDAI-2000 scores with different cut-off scores used to define active disease with SLEDAI-2000.

	Treatment Increased	Treatment Reduced	Treatment Not Changed
BILAG-2004 A or B and SLEDAI-2000 < 10 (n = 37)	18 (48.7%)	4 (10.8%)	15 (40.5%)
BILAG-2004 A or B and SLEDAI-2000 < 9 (n = 37)	18 (48.7%)	4 (10.8%)	15 (40.5%)
BILAG-2004 A or B and SLEDAI-2000 < 8 (n = 35)	17 (48.6%)	3 (8.6%)	15 (42.9%)
BILAG-2004 A or B and SLEDAI-2000 < 7 (n = 35)	17 (48.6%)	3 (8.6%)	15 (42.9%)
BILAG-2004 A or B and SLEDAI-2000 < 6 (n = 35)	17 (48.6%)	3 (8.5%)	15 (42.9%)
BILAG-2004 A or B and SLEDAI-2000 < 5 (n = 30)	15 (50%)	3 (10%)	12 (40%)
BILAG-2004 A or B and SLEDAI-2000 < 4 (n = 16)	9 (56.3%)	0	7 (43.7%)
BILAG-2004 A or B and SLEDAI-2000 < 3 (n = 16)	9 (56.3%)	0	7 (43.7%)
BILAG-2004 A or B and SLEDAI-2000 < 2 (n = 16)	2 (50%)	0	2 (50%)
BILAG-2004 A or B and SLEDAI-2000 = 0 (n = 4)	2 (50%)	0	2 (50%)

Table 3-16: Treatment analysis of patients with low BILAG-2004 and high SLEDAI-2000 scores with different cut-off scores used to define active disease with SLEDAI-2000.

	Treatment Increased	Treatment Reduced	Treatment Not Changed
BILAG-2004 C or D or E and SLEDAI-2000 ≥ 6 (n = 5)	2 (40%)	0	3 (60%)
BILAG-2004 C or D or E and SLEDAI-2000 ≥ 5 (n = 5)	2 (40%)	0	3 (60%)
BILAG-2004 C or D or E and SLEDAI-2000 ≥ 4 (n = 11)	2 (18.2%)	2 (18.2%)	7 (63.6%)
BILAG-2004 C or D or E and SLEDAI-2000 ≥ 3 (n = 11)	2 (18.2%)	2 (18.2%)	7 (63.6%)
BILAG-2004 C or D or E and SLEDAI-2000 ≥ 2 (n = 14)	2 (14.3%)	2 (14.3%)	10 (71.4%)
BILAG-2004 C or D or E and SLEDAI-2000 ≥ 1 (n = 15)	2 (13.3%)	2 (13.3%)	11 (73.3%)

Of those patients with high activity on the BILAG-2004 index but low SLEDAI-2000 scores, 48.6% had their treatment increased. On the other hand, 60% of those with low activity on the BILAG-2004 index but high SLEDAI-2000 scores had their therapy reduced or unchanged. The results were similar when data from the local rater were used (data not shown).

I looked at the effect of using different cut-off scores used to define active disease with SLEDAI-2000 (Tables 3-15 and 3-16). With a lower cut-off SLEDAI-2000 score, the number of patients with high activity on the BILAG-2004 index but low SLEDAI-2000 scores became less, particularly when the cut-off score was 4 or below. However, the proportion of these patients who had their therapy increased remained the same (around 50%). Even at the SLEDAI-2000 score of zero, there were 4 patients (4.3%) with high activity on the BILAG-

2004 index and 2 (50%) of them had their therapy increased. The lowering of the cut-off SLEDAI-2000 scores did not make any difference in the number of patients with low activity on the BILAG-2004 index but high SLEDAI-2000 scores who had their therapy increased. Therefore, it appears that SLEDAI-2000 index is less able to capture active disease requiring increased therapy as compared to the BILAG-2004 index.

3.7 Discussion

These studies had demonstrated that the BILAG-2004 index (with some modifications) was reliable for assessing disease activity in SLE. They were performed in the setting of routine clinical setting that should be close to actual clinical practice. All the raters involved were consultants or senior specialist registrars experienced in the care of patients with SLE. This design was based on the one used in the reliability study of the Classic BILAG index (Hay et al. 1993). However, this study involved more patients (97 versus 50) and more raters (14 versus 7). Furthermore, the agreement and reliability of the index in this study was assessed using several measures (kappa statistic, ICC, level of agreement and major disagreements) instead of depending solely on kappa statistic, which was the case in the study by Hay et al (Hay et al. 1993). This fact is important as there is no single measure that captures all aspects of reliability and agreement.

The main drawback of kappa statistic as an overall measure of performance is its dependence on marginal frequencies which are the row and column totals in the typical table of agreement. Kappa statistic does not perform well when there is a zero total or when there is little discrimination or differentiation within the population under study. This statistic gives low values for reliability, albeit correctly, but says little about level of agreement. This was the case in the Gastrointestinal, Ophthalmic and Constitutional systems as there was

clustering of patients in Grade D or E, resulting in low kappa statistic values when there was a corresponding good level of agreement. Apart from that, the kappa statistic is affected by the number of categories in the measurement scale (Maclure et al. 1987). With more categories, it is more difficult to classify subjects correctly and lower kappa statustic value is the usual result.

The ICC is similarly a measure of reliability designed primarily to assess the ability to distinguish patients from each other. Thus it, like kappa statistic, is affected by the homogeneity of the population studied. With a more heterogeneous population, the value of ICC will be higher. Hence, it is possible to have good agreement with low reliability if the population studied lacks variability (homogeneous). This feature contributes to the drop in the ICC value for the Constitutional system in Exercise 2 when the item fatigue was omitted resulting in a more homogeneous population.

These studies identified four items (fatigue, lupus migraine/cluster/tension headache, mood disorders and anxiety disorder) demonstrating poor agreement between raters and their removal resulted in rather large improvement in agreement and reduction in major disagreements. Their removal seems appropriate as it is difficult to attribute these items to SLE disease activity with certainty and there are specific validated indices available to measure them. Further issues relating to face and content validity of the index were identified during the second reliability exercise. These issues were minor and did not affect the reliability results:

1. Criteria for defining aseptic meningitis were too stringent and it was recommended that this be relaxed to require only those that would be essential for diagnosis (fever, headache and abnormal cerebrospinal fluid).

2. The definition of demyelinating syndrome required the presence of at least one previous event and this would preclude those presenting as the first episode, hence it was recommended that this requirement should be omitted.
3. Definition of cognitive dysfunction required corroborating history from a third party which would be difficult to record if the patient attended alone but was able to give a good account of it. It was therefore recommended that this requirement be omitted.
4. It was noted that tenosynovitis was absent from the index and the recommendation was to include this along with moderate arthritis and tendonitis.
5. The criteria for active nephritis were based on the old WHO Classification and it was recommended that this should incorporate the latest classification which were the WHO Classification 1995 (Churg J et al. 1995) and International Society of Nephrology/Renal Pathology Society (ISN/RPS) Classification 2003 (Weening et al. 2004).
6. The histological criteria for active nephritis included the clause "since previous assessment if seen less than 3 months" which was rather ambiguous and confusing. It was recommended that this clause be removed.
7. The current criteria for proteinuria for Grades A and B which required the proteinuria to rise from more than 1g/day (or equivalent) by more than 50% only took into consideration deteriorating proteinuria and took no account of stable but significant proteinuria. It was recommended that the Grades A and B criteria for proteinuria be changed to that of more than 1g/day (or equivalent) that has not improved/decreased by 50%.
8. The current upper threshold of thrombocytopenia (99 x 10^9/litre) for a score of B in Haematology system was deemed to be too high for consideration of steroid or

immunosuppressive/cytotoxic therapy. Hence it was recommended that this threshold be lowered to 49 x 10^9/litre.

The improvement in agreement from Exercise 1 to Exercise 2 is notable. It is likely that the improvement in the glossary of the index, omission of ambiguous items and training of raters contributed substantially to this improvement. The BILAG-2004 index is a comprehensive index with 97 items and clearly, there will be a need for formal training of users if this index is to be used in the setting of clinical trials.

One of the limitations of the reliability study is the clustering of patients with low level of disease activity, hence the index has not been tested throughout its full range in particular in the Gastrointestinal and Ophthalmic systems which limits the applicability of the results of this study. In fact, the reliability of the Ophthalmic system has not been addressed adequately in this study as the scoring of most of the items in this system requires the assessment by an ophthalmologist. Therefore, the reliability of the Ophthalmic system would be best assessed when the two raters are ophthalmologists. The rarity of these manifestations will make undertaking such a study very difficult to organise. However, these manifestations remain important at an individual level as they do determine the management and outcome.

This study also represents the first reliability study of SLEDAI-2000 index in routine clinical practice and it demonstrates good inter-rater agreement, hence this index is reliable in assessment of SLE disease activity which is reassuring. This is consistent with the results of reliability studies involving the original SLEDAI index (Hawker et al. 1993; Petri et al. 1992; Gladman et al. 1992; Liang et al. 1989). However, the reliability of SLEDAI-2000 index was not as good as would be expected with disagreement in the scores between the two raters in 46% of patients. This is despite training being provided to the raters and the fact that this index is considered to be the least complicated disease activity index.

As both the BILAG-2004 and SLEDAI-2000 indices were used in the second reliability exercise, comparison was made on the ability of these two indices to detect active disease requiring increase in therapy. It appears that SLEDAI-2000 is less able to capture disease activity when compared to the BILAG-2004 index, which is not surprising as it has far fewer items than the BILAG-2004 index (24 versus 97). SLE is a multi-system disease that may affect any system and has varied manifestations within patient and between patients, hence a comprehensive index is required to capture all aspects of active disease.

Unfortunately, SLEDAI-2000 fails to capture several clinically important manifestations of active disease such as peripheral neuropathy, myelopathy, interstitial alveolitis and haemolytic anaemia. Apart from that, some manifestations could not be scored in SLEDAI-2000 as the criteria and definitions set out are too stringent. From the experience of the physicians involved in this study, there was difficulty in scoring for organic brain syndrome, arthritis, pleurisy and pericarditis in SLEDAI-2000 despite patients having these manifestations. For example, to score for pleurisy in SLEDAI-2000, the requirements are that of 'pleuritic chest pain with pleural rub, effusion or pleural thickening'. However, it is not uncommon for SLE patients with pleurisy to present with just pleuritic chest pain in the absence of pleural rub, pleural effusion or pleural thickening. The situation is similar for arthritis, pericarditis and organic brain syndrome.

Other contributing factors include the weighting system used and its inability to distinguish the different severity of manifestations. Furthermore, it is unable to detect improvement or worsening of a manifestation, as this can only be recorded as either absent or present. As an illustration, thrombocytopenia (defined as platelet count less than $100x10^9$ per liter) has a weighted score of one. It does not differentiate between severe thrombocytopenia (platelet count less than $25x10^9$ per liter) and mild thrombocytopenia (platelet count more

than $50x10^9$ per liter). The former would warrant increase in treatment, which is not the case for the latter but both would derive the same score in SLEDAI-2000. Due to the weighted score of one, patients with only severe thrombocytopenia (platelet count less than $25x10^9$ per liter) as its clinical manifestation of disease activity would give a total score of 1 (or 5 if the patient also has low complement and elevated anti-dsDNA antibodies) which falls short of the cut-off of 6 for active disease (Abrahamowicz et al. 1998). Even with a lower cut-off, SLEDAI-2000 index still fails to capture significant numbers of clinically important manifestations of active disease.

In the comparison of the BILAG-2004 and SLEDAI-2000 indices, an overall BILAG-2004 score was used in the analysis as the BILAG-2004 index was developed as an ordinal scale index and not intended for the individual system scores to be summated. It was felt that the best way to represent overall disease activity in any individual patient was to use the highest score achieved by any system within the index. This seems logical as any patient with any system scoring a Grade A or B should be categorised as having active disease regardless of how many systems have a score of A or B, and one system with active disease should be sufficient to determine the level of therapy required to treat the patient, the premise on which the scoring for the index is derived. In fact, using this overall score would create a ceiling effect and may actually put the BILAG-2004 index at a disadvantage from an analysis point of view. For example, a patient with mild mouth ulcers and mild inflammatory arthritis would score Grade C in Mucocutaneous and Musculoskeletal systems, leading to an overall score of Grade C but may have therapy increased with hydroxychloroquine or a small increase in corticosteroids dose.

Increased in therapy was used as the reference standard for active disease in the absence of a good alternative standard. With this as the benchmark, clinically significant

manifestations of active disease (that are being treated) are examined, which makes it very unlikely that the BILAG-2004 index is over-estimating disease activity. Although the BILAG-2004 index is based on the principle of intention to treat, using change in therapy as the reference standard will not bias the analysis in favour of the BILAG-2004 index as actual change in therapy does not determine the scoring (only the presence of active manifestations will influence the scoring). Furthermore, the scoring of the index was not available to the local rater when the treatment decision was made and most investigators were not sufficiently aware of the scoring algorithms. To minimise this possible bias further, the external rater score was used for analysis as the treatment decisions were made by the local rater.

The result of the ability to capture active disease needs to be interpreted with caution as this is a cross sectional study which only provides a snapshot of disease activity at the time of assessment. This does not take into account the level of disease activity prior to the assessment and this may explain the reduction in treatment in patients with high activity, if the current level of activity represents an improvement from a higher previous level of activity such as from Grade A to B in the BILAG-2004 index or SLEDAI-2000 score of 18 to 10. Other unaccounted factors that will influence treatment decision include treatment history (particularly if there has been recent initiation of cytotoxic therapy where further increase in therapy is unlikely) and patients' opinion (such as refusal to increase therapy). One method to overcome this is to use physician's intention of treatment (rather than actual change in treatment) but this may bias the result in favour of the BILAG-2004 index as the physician may record an intention to increase treatment when manifestations are recorded in the BILAG-2004 index.

Further study with larger number of patients is required to determine the optimal SLEDAI-2000 cut-off score for active disease. This assessment is clearly important as it is

used in clinical studies to differentiate patients between the two disease states (active or inactive) and to determine eligibility for inclusion in clinical trials. This issue will be addressed in the cross sectional study.

Several changes were made to the BILAG-2004 index following the results of the 2 reliability exercises and the revised index (Appendix 6) was used in the cross sectional and longitudinal studies.

Chapter 4 Cross Sectional Study

4.1 Abstract

Objectives:

1. To determine if the BILAG-2004 index has construct and criterion validity.
2. To compare the sensitivity and specificity of the BILAG-2004, Classic BILAG and SLEDAI-2000 indices for the assessment of SLE disease activity.

Methods:

SLE patients were recruited in a multi-centre cross sectional study. Data were collected on SLE disease activity (the BILAG-2004, Classic BILAG & SLEDAI-2000 indices), investigations and therapy. Overall BILAG-2004 and Classic BILAG scores were used in the analysis. For construct validity, the constructs used for comparison were ESR, C3, C4, anti-dsDNA and SLEDAI-2000 scores. For criterion validity, increase in therapy was used as the reference standard for active disease. Active disease according to the BILAG-2004 and Classic BILAG indices was defined as overall scores of A or B. Receiver operating characteristic (ROC) curves were used to determine the best cut-off score for SLEDAI-2000 index to define active disease. Statistical analyses were based on logistic regression. Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were calculated.

Results:

A total of 369 SLE patients were recruited contributing 1510 assessments. Increasing overall BILAG-2004 scores were associated with increasing ESR ($p<0.01$), decreasing C3 ($p<0.01$), decreasing C4 ($p<0.01$), elevated anti-dsDNA ($p<0.01$) and increasing SLEDAI-2000 scores

($p<0.01$). Increase in therapy was more frequent with higher overall BILAG-2004 scores. Active disease scores (overall BILAG-2004 scores of A and B) were significantly associated with an increase in therapy (OR 19.3, $p<0.01$). SLEDAI-2000 had lower sensitivity, specificity, PPV and NPV when compared to the BILAG-2004 and Classic BILAG indices. The BILAG-2004 and Classic BILAG indices had comparable sensitivity, specificity, PPV and NPV.

Conclusions:

The BILAG-2004 index has construct and criterion validity. SLEDAI-2000 index does not appear to capture active disease requiring increase in therapy as well as the BILAG-2004 or Classic BILAG indices.

4.2 Objectives

1. To determine if the BILAG-2004 index has construct validity and criterion validity in assessing SLE disease activity.
2. To compare the sensitivity and specificity of the BILAG-2004, Classic BILAG and SLEDAI-2000 indices for the assessment of SLE disease activity.

4.3 Introduction

With the reliability of the BILAG-2004 index established, the next step in the validation process is to determine if this index has construct and criterion validity. Construct validity is demonstrated if this index correlates with other surrogate measures (constructs) of disease activity. This index has criterion validity if it agrees with the gold standard. As there is no gold standard for SLE disease activity, change in therapy is used as the reference standard instead. A cross sectional study was designed specifically for this purpose and the results are presented in this chapter.

The SLEDAI-2000 index produces a global numerical score but there is debate about the appropriate cut-off score as an indicator of active disease. This is an important issue as it is used in clinical studies to differentiate between two disease states (active or inactive) and more so, in clinical trials where it is used to determine eligibility. There have been very few studies to define the appropriate cut-off score and currently available data have been conflicting with one study suggesting a cut-off score of 6 (Abrahamowicz et al. 1998) while two others suggested a lower cut-off of 4 (Yee et al. 2007b; Gladman et al. 2000b). In addition, all these studies were hampered by either small sample size, abstracted case histories, retrospective assessments or limited statistical analysis. This issue is addressed in

this cross sectional study by determining the optimal cut-off score for SLEDAI-2000 to define active disease requiring increase in therapy. Furthermore, the performances of the BILAG-2004, Classic BILAG and SLEDAI-2000 indices were compared with regards to sensitivity, specificity, positive predictive value and negative predictive value in assessment of SLE disease activity.

4.4 Methods

This was a multi-centre cross sectional study involving 8 centres in the United Kingdom. This study commenced in March 2005 and was completed in August 2006. At every assessment, data on disease activity, investigations and treatment were collected. Disease activity was assessed using the BILAG-2004, Classic BILAG and SLEDAI-2000 indices. All clinicians involved in this study had been trained to use all three disease activity indices. The majority of patients recruited into this study had more than one assessment during this period.

4.5 Analysis

Overall BILAG-2004 and overall Classic BILAG scores were used in the analysis. These overall scores were determined by the highest score achieved by any system in the respective indices.

Construct Validity

The constructs used in this validation study were erythrocyte sedimentation rate (ESR), complements C3 and C4, anti-dsDNA antibody levels and SLEDAI-2000 score. It was expected that the BILAG-2004 index would have a positive correlation or association with

ESR, anti-dsDNA antibody and SLEDAI-2000 score (as they increase with disease activity), and negative correlation or association with complement C3 and C4 levels (as they fall with disease activity).

The levels of ESR, anti-dsDNA antibody, C3 and C4 were determined locally at the respective participating centres. As the laboratory kits used were not the same in all the centres, the normal values for anti-dsDNA antibody, C3 and C4 differed between centres. Therefore, for the purpose of the analysis, all these constructs were divided into ordinal categories as follows:

(a) ESR – normal (0 – 30), elevated (31 – 60) and markedly elevated (> 60)

(b) C3 and C4 – normal, low and very low (≤ half of lower normal limit)

(c) Anti-dsDNA – normal, elevated and very high (> 5 times upper normal limit)

(d) SLEDAI-2000 scores – inactive (0), mild activity (1 – 3), active (4 – 12) and very active (> 12)

Maximum-likelihood ordinal logistic regression (with robust variance estimation) was used to assess construct validity with overall BILAG-2004 score as the outcome variable and the constructs as the explanatory variable. The normal or inactive category of the constructs was used as the baseline comparator for the other categories. Results were reported in terms of odds ratio (OR) with a 95% confidence interval (CI).

Criterion Validity

Change in therapy was used as the reference standard or criterion for this analysis. In this study, change in therapy was the change in treatment following assessment, which was the difference in treatment after the patient was assessed, as compared to the treatment that the patient was on prior to the assessment. The definition of change in treatment has been

described in Chapter 2. Change in therapy was classified into two categories, namely 'increase in therapy' and 'no increase in therapy'.

For this analysis, active disease according to the BILAG-2004 and Classic BILAG indices was defined as overall scores of A or B. For SLEDAI-2000, receiver operating characteristic (ROC) curve analysis was used to determine the optimal cut-off score for active disease associated with a change in therapy. This cut-off score was then used for comparison with the other indices.

Maximum likelihood logistic regression (with robust variance estimation) was used with change in therapy as the outcome variable and the respective disease activity index score as the explanatory variable. The results were reported in terms of odds ratios (OR) with 95% confidence intervals (CI). Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were calculated with the same methodology, but applied only to appropriate subsets of patients and with only the intercept term estimated. For calculation of sensitivity and its 95% confidence interval, only observations which recorded an increase in treatment were used. Specificity, PPV and NPV were calculated using only patients with no increase in treatment, active disease and minimally active disease, respectively.

4.6 Results

A total of 369 SLE patients were recruited and they contributed 1510 assessments for the analysis. The demographics of the patients are summarised in Table 4-1. The majority (88.6%) of patients had more than one assessment during the study period. The distribution of disease activity according to the BILAG-2004, Classic BILAG and SLEDAI-2000 indices is summarised in Table 4-2.

Table 4-1: Demographics of patients recruited into the cross sectional study (n=369).

Patient Characteristics	
Female sex (%)	92.7
Mean age in years (SD)	41.6 (13.2)
Ethnicity (%)	
Caucasian	59.9
Afro-Caribbean	18.4
South Asian	18.4
Oriental	1.4
Others	1.9
Mean disease duration in years (SD)	8.8 (7.7)
Number of assessments (%)	
1	11.4
2	12.5
3	19.8
4	18.7
5	15.2
6	10.0
7 or more	8.1

Table 4-2: Distribution of disease activity based according to the BILAG-2004, Classic BILAG and SLEDAI-2000 indices (n=1510).

Disease Activity Indices	Number (%)
Overall BILAG-2004 score	
A	98 (6.5)
B	390 (25.8)
C	721 (47.8)
D/E	301 (19.9)
Overall Classic BILAG score	
A	95 (6.3)
B	435 (28.8)
C	834 (55.2)
D/E	146 (9.7)
SLEDAI-2000 score	
More than 12	28 (1.9)
4 to 12	539 (35.7)
1 to 3	419 (27.8)
0	524 (34.7)

4.6.1 Construct Validity

The distribution of disease activity according to the BILAG-2004 index and constructs (cross tabulated against disease activity) is summarised in Table 4-3.

Table 4-3: Cross tabulation of overall BILAG-2004 scores with constructs (ESR, anti-dsDNA antibody, C3, C4 and SLEDAI-2000 score).

Constructs	Overall BILAG-2004 Scores			
	A	**B**	**C**	**D**
ESR (n=787)				
Normal (0 – 30)	29	158	267	100
Elevated (31 – 60)	23	60	70	15
Markedly elevated (> 60)	13	25	20	7
Anti-dsDNA (n=1413)				
Normal	49	228	506	216
Elevated	25	100	119	66
Very high	15	38	42	9
C3 (n=1463)				
Normal	61	279	579	270
Low	29	100	106	25
Very Low	4	4	6	0
C4 (n=1366)				
Normal	44	218	467	243
Low	18	76	132	37
Very Low	25	49	51	6
SLEDAI-2000 (n=1510)				
Inactive (0)	2	22	306	194
Mild activity (1 – 3)	4	93	240	82
Active (4 – 12)	75	264	175	25
Very active (> 12)	17	11	0	0

Construct - ESR

ESR values were available for 787 (52.1%) assessments. There was a significant association between increasing ESR and higher overall BILAG-2004 scores (Table 4-4). The 2 degrees of freedom test for an association between overall BILAG-2004 score and ESR was

statistically significant ($p<0.001$). When ESR was analysed as a continuous variable, the result was similar ($p=0.0002$). When only the first visit from each patient was used in the analysis, the result was similar (Table 4-5).

Construct - Anti-dsDNA antibody

Anti-dsDNA antibody levels were available for 1413 (93.6%) assessments. Increasing levels of anti-dsDNA were significantly associated with higher overall BILAG-2004 scores (Table 4-4). The 2 degrees of freedom test for an association between overall BILAG-2004 score and anti-dsDNA was statistically significant ($p=0.0008$). When only the first visit from each patient was used in the analysis, the result was similar (Table 4-5).

Construct - C3 and C4

C3 and C4 levels were available in 1463 (96.9%) and 1366 (90.5%) assessments respectively. There was significant association between lower C3 and C4 levels with higher overall BILAG-2004 scores (Table 4-4). For both the models, the 2 degrees of freedom test were statistically significant ($p<0.0001$). When only the first visit from each patient was used in the analysis, the result was similar (Table 4-5).

Construct - SLEDAI-2000

SLEDAI-2000 scores were available for all the assessments. Higher SLEDAI-2000 scores were significantly associated with higher overall BILAG-2004 scores (Table 4-4). The 3 degrees of freedom test for an association between overall BILAG-2004 score and SLEDAI-2000 score was statistically significant ($p<0.001$). The results were similar when

SLEDAI-2000 score was analysed as a continuous variable ($p<0.0001$). When only the first visit from each patient was used in the analysis, the result was similar (Table 4-5).

Table 4-4: Association of ESR, anti-dsDNA antibody, C3, C4 and SLEDAI-2000 scores respectively with higher overall BILAG-2004 scores.

Constructs	Odds Ratio (95% Confidence Interval)
ESR	
Normal	1.0
Elevated (31 – 60)	2.1 (1.4, 3.0)
Markedly elevated (> 60)	2.9 (1.4, 6.2)
Anti-dsDNA Antibody	
Normal	1.0
Elevated	1.5 (0.99, 2.1)
Very high (> 5 times upper normal limit)	2.7 (1.6, 4.8)
C3	
Normal	1.0
Low	2.5 (1.8, 3.5)
Very Low (≤ half of lower normal limit)	5.0 (1.6, 15.5)
C4	
Normal	1.0
Low	1.7 (1.2, 2.3)
Very Low (≤ half of lower normal limit)	4.2 (2.5, 6.9)
SLEDAI-2000	
No activity (0)	1.0
Mild activity (1 – 3)	3.0 (2.1, 4.4)
Active (4 – 12)	20.0 (13.6, 29.5)
Very active (> 12)	232.7 (108.4, 499.2)

Table 4-5: Association of ESR, anti-dsDNA antibody, C3, C4 and SLEDAI-2000 scores respectively with higher overall BILAG-2004 scores (when only first visit from each patient was used in the analysis).

Constructs	Odds Ratio (95% Confidence Interval)
ESR	
Normal	1.0
Elevated (31 – 60)	2.9 (1.6, 5.2)
Markedly elevated (> 60)	5.7 (2.1, 15.4)
Anti-dsDNA Antibody	
Normal	1.0
Elevated	1.4 (0.9, 2.2)
Very high (> 5 times upper normal limit)	2.5 (1.1, 5.5)
C3	
Normal	1.0
Low	2.2 (1.3, 3.6)
Very Low (≤ half of lower normal limit)	27.8 (2.6, 295.3)
C4	
Normal	1.0
Low	2.3 (1.3, 4.0)
Very Low (≤ half of lower normal limit)	3.3 (1.6, 6.6)
SLEDAI-2000	
No activity (0)	1.0
Mild activity (1 – 3)	3.8 (2.2, 6.4)
Active (4 – 12)	24.6 (13.7, 44.5)
Very active (> 12)	1386.3 (160.7, 11955.9)

4.6.2 Criterion Validity

Of the 1510 assessments, 342 assessments (22.6%) resulted in an increase in therapy while 320 assessments (21.2%) had treatment reduction and 848 assessments (56.2%) were not followed by a change in therapy (Table 4-6). There was increasing likelihood of an increase in therapy with higher overall BILAG-2004 score (Table 4-7). When only the first

visit from each patient was used in the analysis, the result was similar but the strength of association (odds ratio) was markedly inflated due the absence of any increase in therapy for overall BILAG-2004 score of D (which was the baseline comparator for the other Grades) (Table 4-8).

Table 4-6: Cross tabulation of overall BILAG-2004 scores with change in therapy (n=1510).

Overall BILAG-2004 Score	Change in Therapy		
	Decrease	No Change	Increase
D	83	213	4
C	197	464	61
B	38	147	205
A	2	24	72

Table 4-7: Association of increase in therapy with overall BILAG-2004 scores.

Overall BILAG-2004 Scores	Odds Ratio (95% Confidence Interval)
A	204.9 (65.6, 639.9)
B	82.0 (29.9, 225.0)
C	6.8 (2.4, 19.1)
D	1.0

Table 4-8: Association of increase in therapy with overall BILAG-2004 scores (when only the first visit from each patient was used in the analysis).

Overall BILAG-2004 Scores	Odds Ratio (95% Confidence Interval)
A	4.79×10^8 *
B	1.4×10^8 (5.3×10^7, 3.7×10^8)
C	1.3×10^7 (4.8×10^6, 3.7×10^7)
D	1.0

* confidence interval could not be estimated

4.6.3 Optimal SLEDAI-2000 Cut-Off Score for Active Disease

ROC curves analysis for SLEDAI-2000 as a predictor of increase in therapy are summarised by the sensitivities, specificities and area under the ROC curve (AUC) in Table 4-9 where PPV and NPV values were also provided. The optimal cut-off score for active disease is 3 or 4 as both had very similar performance with the best area under the curve.

Table 4-9: Receiver operating characteristic curves analysis of SLEDAI-2000 index.

Cut-off Score	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	AUC
2	87.7% (83.2, 91.1)	43.0% (38.5, 47.6)	31.1% (27.9, 34.4)	92.3% (89.5, 94.3)	0.65
3	71.9% (65.8, 77.3)	70.7% (66.1, 75.0)	41.8% (37.5, 46.3)	89.6% (87.2, 91.6)	0.71
4	70.8% (64.6, 76.3)	72.2% (67.5, 76.4)	42.7% (38.2, 47.2)	89.3% (87.0, 91.4)	0.71
5	44.4% (38.0, 51.1)	87.5% (84.3, 90.1)	51.0% (45.0, 57.0)	84.3% (81.9, 86.5)	0.66
6	42.1% (35.9, 48.6)	88.1% (84.9, 90.7)	50.9% (44.7, 57.0)	83.9% (81.4, 86.0)	0.65
7	28.4% (22.4, 35.1)	93.3% (90.7, 95.2)	55.4% (48.0, 62.6)	81.6% (79.0, 84.0)	0.61
8	26.3% (20.6, 32.9)	93.8% (91.2, 95.7)	55.6% (47.6, 63.2)	81.3% (78.7, 83.6)	0.61
9	17.3% (13.1, 22.4)	96.7% (95.1, 97.9)	60.8% (52.0, 69.0)	80.0% (77.4, 82.3)	0.57
10	17.3% (13.1, 22.4)	97.0% (95.3, 98.1)	62.8% (53.5, 71.1)	80.0% (77.5, 82.3)	0.57

The ROC curve analysis was repeated using a modified SLEDAI-2000 score that excluded the items anti-dsDNA antibody and complements (Table 4-10). The optimal cut-off

score for active disease using this modified score was 2. There was improvement in the performance of SLEDAI-2000 index with the exclusion of these two items.

Table 4-10: Receiver operating characteristic curves analysis of SLEDAI-2000 index using modified total score excluding the items anti-dsDNA antibody and complements.

Cut-off Score	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	AUC
1	82.7% (77.7, 86.9)	69.6% (65.6, 73.4)	44.4% (40.4, 48.4)	93.2% (91.1, 94.9)	0.76
2	81.6% (76.5, 85.7)	72.6% (68.7, 76.2)	46.6% (42.5, 50.7)	93.1% (91.0, 94.7)	0.77
3	57.6% (51.2, 63.8)	86.0% (82.6, 88.7)	54.6% (49.0, 60.0)	87.4% (85.0, 89.4)	0.72
4	54.1% (47.6, 60.4)	86.9% (83.6, 89.7)	54.7% (58.9, 60.4)	86.6% (84.2, 88.7)	0.71
5	29.8% (24.5, 35.7)	95.2% (93.3, 96.6)	64.6% (57.1, 71.4)	82.2% (79.9, 84.4)	0.63
6	28.9% (23.7, 34.9)	95.5% (93.6, 96.8)	65.1% (57.4, 72.1)	82.1% (79.7, 84.3)	0.62
7	17.5% (13.2, 23.0)	97.4% (96.0, 98.4)	66.7% (57.4, 74.8)	80.1% (77.6, 82.5)	0.57
8	16.7% (12.4, 22.0)	97.9% (96.5, 98.8)	70.4% (60.1, 78.9)	80.1% (77.5, 82.4)	0.57
9	10.5% (7.5, 14.5)	99.3% (98.7, 99.7)	81.8% (69.4, 89.9)	79.1% (76.6, 81.4)	0.55

4.6.4 Comparison of Indices

For this analysis, the SLEDAI-2000 cut-off score of 3 was used. Active disease scores according to the BILAG-2004 (OR 19.3, 95% CI: 14.1, 26.4), Classic BILAG (OR 19.3, 95% CI: 14.0, 26.7) and SLEDAI-2000 (OR 6.2, 95% CI: 4.6, 8.3) indices were significantly

associated with an increase in therapy. The sensitivity, specificity, PPV and NPV of the three disease activity indices are summarised in Table 4-11.

The BILAG-2004 and Classic BILAG indices had equivalent sensitivity, specificity, PPV and NPV. Consistent with its lower estimated odds ratio, it was apparent that SLEDAI-2000 had lower sensitivity, specificity, PPV and NPV compared to the BILAG-2004 and Classic BILAG indices. The results were not different when SLEDAI-2000 cut-off score of 4 was used instead of 3. Even when the modified SLEDAI-2000 score that excluded the items anti-dsDNA antibody and complements was used (with a cut-off score of 2), the improved performance was still not as good as the BILAG-2004 or Classic BILAG indices.

Table 4-11: Sensitivity, specificity, positive predictive value and negative predictive value for increase in therapy of the BILAG-2004, Classic BILAG and SLEDAI-2000 indices.

	BILAG-2004 (95% CI)	**Classic BILAG (95% CI)**	**SLEDAI-2000 (95% CI)**
Sensitivity (%)	81.0 (76.4, 84.9)	83.6 (79.1, 87.3)	71.9 (65.8, 77.3)
Specificity (%)	81.9 (78.4, 85.0)	79.1 (75.5, 82.3)	70.7 (66.1, 75.0)
PPV (%)	56.8 (51.7, 61.7)	54.0 (58.6, 49.3)	41.8 (37.5, 46.3)
NPV (%)	93.6 (91.9, 95.0)	94.3 (92.6, 95.6)	89.6 (87.2, 91.6)

4.7 Discussion

The results of this large multi-centre cross sectional study demonstrated that the BILAG-2004 index was a valid measure of SLE disease activity as it had construct and criterion validity. Its construct validity was confirmed with the expected association between this index and ESR, complement C3, complement C4, anti-dsDNA antibody and SLEDAI-2000 score. The association was strongest with SLEDAI-2000 score as there was a significant number of assessments with active disease that did not have high ESR, low C3, low C4 or high anti-dsDNA antibody (Table 3-4). The criterion validity of the BILAG-2004 index was

confirmed with increasing strength of association between increasing overall BILAG-2004 score and increase in therapy.

This study also demonstrated that the optimal SLEDAI-2000 cut-off score for definition of active disease linked to the need to increase therapy was 3 or 4. This was consistent with the results from our reliability study (Yee et al. 2007a) and another study by Gladman et al which used SLEDAI index rather than SLEDAI-2000 index (Gladman et al. 2000b). However, this is different from the cut-off of 6 that was suggested by Abrahamowicz et al (Abrahamowicz et al. 1998). This is most likely related to the difference in the study design. This was a prospective study derived from a large number of patients within actual clinical practice whereas the earlier study was based on hypothetical situations derived from 30 abstracted case histories. As such, the result of this study is highly relevant and applicable to clinical practice.

It is not surprising that SLEDAI-2000 index does not capture disease activity requiring increase in therapy as well as the BILAG-2004 and Classic BILAG indices. The latter two indices are more comprehensive (97 and 86 items respectively) as compared to SLEDAI-2000 index which has only 24 items. This property (comprehensiveness) is essential to capture all aspects of disease activity in a multi-system disease with diverse manifestations.

One of the drawbacks of SLEDAI-2000 index is the inclusion of anti-dsDNA antibody and complement levels. Studies have shown that single estimation of these markers do not reflect SLE disease activity well, longitudinal changes in the levels of these markers are much more helpful particularly in lupus nephritis (Ho et al. 2001a; Ho et al. 2001b; Esdaile et al. 1996; Ter Borg et al. 1990; Swaak et al. 1986; Swaak et al. 1982). Furthermore, they are not abnormal in a significant proportion of patients with active disease and it is not uncommon for them to be abnormal in the absence of disease activity as shown in this study and many other

studies (Ng et al. 2006; Gladman et al. 2003; Walz LeBlanc et al. 1994; Petri et al. 1991). This is reflected in the analysis of this study as the performance of SLEDAI-2000 did improve when these two items (anti-dsDNA antibody and complements) were removed. Another contributing factor is the weighting system used in SLEDAI-2000 as there appears to be excessive weighting for some minor manifestations such as alopecia and mouth ulcers that are not always treated. Furthermore, SLEDAI-2000 provides less scope for differentiating the severity of manifestations of active disease or indicating any change in the status of disease manifestation (improving, worsening or remaining unchanged).

In this study, no difference in the performance between the BILAG-2004 and Classic BILAG indices was demonstrated. This was not unexpected as the main difference between the two indices was the addition of ophthalmic and gastrointestinal manifestations in the BILAG-2004 index. Active disease manifestations in these two systems were not common in this study; there were only 6 assessments (from 5 patients) with Grades A or B score in the Gastrointestinal system and 8 assessments (from 3 patients) with Grades A or B score in the Ophthalmic system. Even though these manifestations are uncommon, they need to be captured as they are important for individual patients in view of the significant morbidity and mortality associated with them.

Overall BILAG-2004 and Classic BILAG scores were used in the analysis as both of these indices were developed as ordinal scale index and were not intended for the individual system scores to be summated into a global score. The best way to represent overall disease activity (as a yes/no dichotomy) for this analysis in any individual patient was to use the highest score achieved by any system within the index. This is logical as any patient with any system scoring a Grade A or B should be categorised as having active disease (requiring therapy in principle) regardless of how many systems have a score of A or B. From an

analysis viewpoint, this may put both these indices at a disadvantage as there is a ceiling effect and this may underestimate the severity of the illness. For example, a patient with 5 systems scoring Grade B is equivalent to a patient with just one system scoring Grade B, with this analysis. However, for clinical trials and outcome studies, it may be more appropriate to consider the number of systems with a given categorical score in the analysis.

Change in therapy was chosen as the reference standard for disease activity in the absence of a better alternative and because it has an operational interpretation. To date, the best benchmark to define active disease is the decision as to whether the disease activity should be treated. A robust definition of change in therapy was used in this study to account for several circumstances whereby therapy was increased for the reasons other than that of active disease, and this will minimise the occurrence of misclassification of change in therapy. Physician's global assessment (PGA) had been used previously as a reference standard but this had been shown to have unsatisfactory performance with poor agreement between physicians in several studies (Wollaston et al. 2004; Gladman et al. 1994; Guzman et al. 1992; Gladman et al. 1992).

Although the BILAG-2004 and Classic BILAG indices were developed based on the principle of the physician's intention to treat, using change in therapy as the reference standard should not explicitly bias the analysis in favour of these two indices as change in therapy does not determine the scoring of the index. Only the presence of manifestations of active disease will influence the scoring. Furthermore, the scoring of these two indices were not available to the physician when the treatment decision was made and it would be difficult to work out the scoring for both of these indices in routine clinical practice without the appropriate reference documentations.

One of the limitations of this study is inherent in the cross sectional design in that it only accounts for disease activity at the time of assessment. This does not take into account the level of disease activity prior to the assessment which will influence the treatment decision. The treatment decision regarding a patient with active disease is very different if prior disease activity was low (such as a change from Grade D to B) when treatment would be increased, as compared to the situation where prior disease activity was high (such as a change from Grade A to B) when there would not be an increase in therapy (in fact there might be a reduction in therapy). Treatment decision is a complex process which takes into account several factors apart from the physician's intention of treatment. This process includes consideration of current therapy, previous therapy (and its response), the patient's opinion (in particular refusal to change therapy as advised) and the presence of co-morbid conditions. Unfortunately, it is not possible to model all these factors into the analysis of this study. This may explain the relatively low positive predictive value of the BILAG-2004 and Classic BILAG indices. On the other hand, the high negative predictive value is reassuring as this indicates that increase in therapy is very unlikely in the absence of high disease activity as measured by the BILAG-2004 and Classic BILAG indices.

In conclusion, the BILAG-2004 index is a valid measure of SLE disease activity. It is more comprehensive, incorporates up-to-date terminology and has a clearer glossary of definitions than the Classic BILAG index. This index also appears to reflect the need to treat SLE disease activity better than SLEDAI-2000 index. This study also addressed the use of BILAG-2004, Classic BILAG and SLEDAI-2000 indices to define active disease at a single time point. It did not address the issue of minimal clinically important change in the scores that is relevant to longitudinal studies. This issue will be addressed in the longitudinal study.

Chapter 5 Face and Content Validity of Renal System

5.1 Abstract

Objectives:

Some changes to the scoring scheme of BILAG-2004 renal system were proposed and this study was to assess the face and content validity of these changes.

Methods:

Data from the prospective multi-centre longitudinal study designed to assess sensitivity to change of the BILAG-2004 index were used for this analysis. At every assessment, data were collected on disease activity (with the BILAG-2004 index), investigations and therapy. Assessments with renal systems score of A, B or C were selected for analysis. Only those assessments with urine protein estimation available were used. Assessments where the renal system score was lower than other system scores were excluded. The change in therapy following these assessments was analysed.

Results:

A total of 353 SLE patients were recruited and there were 1771 assessments. Of these, 181 (10.2%) assessments from 45 patients had renal system score of A, B or C. 130 assessments from 36 patients had urine protein estimation and the renal system score was not lower than that of any of the other systems. 70 assessments from 26 patients had no other renal factors influencing the renal system score. The proposed changes to BILAG-2004 renal system scoring did improve the performance of the index when the treatment decision was analysed.

Conclusions:

Changes to the scoring scheme of BILAG-2004 renal system were valid.

5.2 Objective

To assess the face and content validity of the proposed changes to the scoring scheme for the renal system of the BILAG-2004 index.

5.3 Introduction

During the course of the longitudinal study (for sensitivity to change), some issues related to the scoring of the Renal system of the BILAG-2004 index based on the level of proteinuria for Grades A and B were noted (please refer to Renal System scoring scheme in Appendix 6). The issues identified were:

1. Over-scoring of the Renal system due to over-reliance on urine protein dipstick despite the presence of other methods of urine protein estimation (24-hours urine protein estimation, urine albumin-creatinine ratio and urine protein-creatinine ratio).
2. The current threshold for improvement (at least 50%) in proteinuria may be too high (especially if the assessments are done 4-weekly) resulting in over-scoring of the Renal system which is not consistent with treatment decisions.
3. The current scoring scheme is weighted towards heavy proteinuria (more than 1g/day or equivalent) and less severe but significant proteinuria (0.5 to 1g/day or equivalent) would score inappropriately low as mild activity (Grade C).

Changes to the scoring of the Renal system for Grades A and B were suggested to take account of the issues raised. These changes were:

1. Urine protein dipstick result would be superseded by other methods of urine protein estimation (urine albumin-creatinine ratio, urine protein-creatinine ratio or 24-hours urine protein) where available.

2. A lower threshold of 25% and 33% for definition of improvement in proteinuria may be more appropriate
3. An additional Grade B criterion of urine protein excretion of at least 0.5g/day (or equivalent)

The dataset from the longitudinal study (sensitivity to change study) was used to assess the validity of these proposed changes to the Renal system scoring scheme.

5.4 Methods

Data from the longitudinal study were used to study the effects of these changes. This was a multi-centre study designed to assess sensitivity to change of the BILAG-2004 index. Patients were followed up prospectively and data were collected for all consecutive visits or encounters (inpatient or outpatient) that the patients had with the physician. Data were collected on disease activity (as assessed using the BILAG-2004 index), investigations and treatment. This study commenced in March 2005 and was completed in April 2007.

5.5 Analysis

Assessments with Renal system score of Grades A, B or C were selected for this analysis, as the proposed changes would have no effect on existing Renal system score of D or E (inactivity). Only those assessments with urine protein estimation (24-hours urine protein estimation, urine albumin-creatinine ratio or urine protein-creatinine ratio) available were used. The change in therapy following these assessments were analysed, which is the difference in the treatment the patient was on after the assessment as compared to the treatment the patient was on prior to the assessment (similar to that which was used in the

cross sectional study). This differed from the definition of change in therapy used in the sensitivity to change analysis whereby it was the difference in treatment between two consecutive visits.

Assessments where the Renal system score was lower than that of other system scores were excluded as the Renal score would not have been influential on the treatment decision, for example an assessment with Renal system score of B and Mucocutaneous system score of A would be excluded as the Mucocutaneous system score would be the main determinant of treatment decision rather than the Renal system score. On the other hand, an assessment with Renal system score of B and Mucocutaneous system score of B would be included as the Renal system score would be a major determinant of the decision to treat.

The aims of these changes were (in order of importance):

1. Reduction in the number/proportion of Renal system Grade A that is associated with decrease in therapy
2. Reduction in the number/proportion of Renal system Grade B that is associated with decrease in therapy, although decrease in therapy was possible if this followed a reduction in activity from Grade A
3. Increase in the number/proportion of Renal system Grade B that is associated with increase in therapy
4. No increase in the number/proportion of Renal system Grade C that is associated with increase in therapy, but increase in therapy would be possible if any other system had a higher score as a result of down-grading of the Renal system score due to the modification

5.6 Results

There was a total of 1767 assessments from 353 patients (the demographics of these patients will be presented in the next chapter on the sensitivity to change study). Out of this total, 181 assessments (10.2%) had Renal system score of A, B or C and these were derived from 45 patients. Of these, 130 assessments (from 36 patients) that had urine protein estimation and for which the Renal system score was not lower than any of the other systems, were used for this analysis. Seventy assessments (from 26 patients) had no other renal factors influencing the Renal system score, such as histology of active nephritis, presence of active urinary sediments, deteriorating renal function or presence of nephrotic syndrome.

With the current Renal system scoring scheme, the treatment for these patients is summarised in Table 5-1.

Table 5-1: Cross tabulation of BILAG-2004 Renal system score (excluding Categories D and E) with change in therapy following assessment.

Renal System Score	Change in Therapy			Total
	No change	Decrease	Increase	
A	11 (30.6%)	2 (5.6%)	23 (63.9%)	36
B	38 (56.7%)	13 (19.4%)	16 (23.9%)	67
C	17 (63.0%)	9 (33.3%)	1 (3.7%)	27
Total	66 (50.8%)	24 (18.4%)	40 (30.8%)	130

Firstly, the effect of urine protein dipstick result not contributing to the Renal system score in the presence of other methods of urine protein estimation (24-hours urine protein, urine albumin-creatinine ratio or urine protein-creatinine ratio) was assessed. The result is summarised in Table 5-2. There was some shift from higher scores (A & B) to lower scores (C & D) and there was a reduction in Renal system score of B that was associated with

decrease in therapy, although there was slight increase in renal system scores of C with treatment increased.

Table 5-2: Cross tabulation of modified BILAG-2004 Renal system score (in which urine protein dipstick result was superseded by other methods of urine protein estimation) with change in therapy following assessment.

Renal System Score	Change in Therapy			Total
	No change	Decrease	Increase	
A	11 (32.4%)	2 (5.9%)	21 (61.8%)	34
B	32 (55.2%)	10 (17.2%)	16 (27.6%)	58
C	21 (58.3%)	12 (33.3%)	3 (8.3%)	36
D	2 (100%)	0	0	2
Total	66 (50.8%)	24 (18.4%)	30 (30.8%)	130

Next, the effect of changing the threshold for definition of improvement in proteinuria to 33% and 25% respectively was considered. This was in addition to the criterion whereby urine protein dipstick result would be superseded by other methods of urine protein estimation. The results are summarised in Tables 5-3 and 5-4. The 25% threshold for definition of improvement in proteinuria performed the best when compared to 33% and 50%, with reduction in Renal system scores of A and B in which there was decrease in therapy.

Table 5-3: Cross tabulation of modified BILAG-2004 Renal system score (in which the threshold for definition of improvement in proteinuria was changed to 33% and urine protein dipstick result was superseded by other methods of urine protein estimation) with change in therapy following assessment.

Renal System Score	Change in Therapy			Total
	No change	Decrease	Increase	
A	9 (29.0%)	2 (6.5%)	20 (64.5%)	31
B	31 (56.4%)	9 (16.4%)	15 (27.3%)	55
C	24 (57.1%)	13 (31.0%)	5 (11.9%)	42
D	2 (100%)	0	0	2
Total	66 (50.8%)	24 (18.4%)	30 (30.8%)	130

Table 5-4: Cross tabulation of modified BILAG-2004 Renal system score (in which the threshold for definition of improvement in proteinuria was changed to 25% and urine protein dipstick result was superseded by other methods of urine protein estimation) with change in therapy following assessment.

Renal System Score	Change in Therapy			Total
	No change	Decrease	Increase	
A	9 (30.0%)	1 (3.3%)	20 (66.7%)	30
B	29 (54.7%)	9 (17.0%)	15 (28.3%)	53
C	26 (57.8%)	14 (31.1%)	5 (11.1%)	45
D	2 (100%)	0	0	2
Total	66 (50.8%)	24 (18.4%)	30 (30.8%)	130

The effect of an additional Renal system Grade B criterion of urine protein excretion of at least 0.5g/day (or equivalent) that has not improved/decreased by 25% was assessed subsequently (Table 5-5). This additional criterion did not appear to have much effect on the performance but it did increase the number of Renal system score of B with reduction in therapy.

Table 5-5: Cross tabulation of modified BILAG-2004 Renal system score (in which the threshold for definition of improvement in proteinuria was changed to 25%, additional Grade B criterion of urine protein excretion of at least 0.5g/day (or equivalent) that has not improved/decreased by 25%, and urine protein dipstick result would be superseded by other methods of urine protein estimation) with change in therapy following assessment.

Renal System Score	Change in Therapy			Total
	No change	Decrease	Increase	
A	9 (30.0%)	1 (3.3%)	20 (66.7%)	30
B	33 (53.2%)	13 (21.0%)	16 (25.8%)	62
C	22 (61.1%)	10 (27.8%)	4 (11.1%)	36
D	2 (100%)	0	0	2
Total	66 (50.8%)	24 (18.4%)	30 (30.8%)	130

The results of the above analysis did support the changes that were proposed.

5.7 Discussion

During the course of the longitudinal study to assess sensitivity to change of the index, some issues with the Renal scoring system of the BILAG-2004 index came to light and changes were suggested. As these changes would not have made any impact on the way the data were collected with the index, the validity of these changes was analysed using this dataset by looking at the way patients with disease activity in the Renal system were treated, following each assessment.

The first of these changes was that urine protein dipstick result would be superseded by other methods of urine protein estimation. This alteration is appropriate as urine protein dipstick measure is a rather crude and inaccurate method of urine protein estimation with a tendency to high false positive results which can lead to over-scoring of the Renal system. For

example, an increase of urine protein dipstick from 1+ to 2+ (equivalent to a Grade B score) with concomitant urine protein excretion increasing from 0.30g/day to 0.45g/day (equivalent to a Grade C score) will result in a change in Renal system score from C to B indicating an increase in activity, but in practice this change would not constitute a significant change requiring an increase in therapy. This was confirmed in the first part of the analysis which showed that this change resulted in a shift from higher scores (A or B) to lower scores (C or D), leading to a reduction in Renal system score of B with decrease in therapy.

The second issue was that the threshold for improvement in proteinuria of at least 50% was considered to be too high. The dataset was used to look at the effect of changing the threshold to 25% and 33% respectively. Although the number of relevant assessments in the analysis were small, the reduction in the threshold did appear to improve the performance of the index with a reduction in the number of assessments with Renal system scores of A and B that had treatment reduced. The threshold for improvement of 25% was considered to have performed the best of the three thresholds tested and was felt to be the most appropriate when this was discussed in a consensus group reviewing the data, involving members of BILAG and 2 nephrologists experienced in the management lupus nephritis (Dr Dwomoa Adu and Dr David Jayne).

It was also noted that the Renal scoring scheme was weighted towards high proteinuria (at least 1g/day or equivalent) and less severe but significant proteinuria (between 0.5 to 1g/day or equivalent) would not be appropriately captured as it would only score Grade C. The additional criterion recommended was to capture this significant proteinuria (between 0.5 to 1g per day or equivalent) as a Grade B score, instead of Grade C. However, this change did not have much effect in the analysis. This was not unexpected as this additional criterion would only have an effect when there was no other renal factor (such as histology of active

nephritis, presence of active urinary sediments and deteriorating renal function) present as these other factors would have resulted in a score of A or B. Therefore, it would only have an effect in less than 40% of the assessments (70 assessments from 26 patients) when there was no other renal factor (histology of active nephritis, presence of active urinary sediments and deteriorating renal function) influencing the renal system score. Even though the data did not strongly support the inclusion of this additional Grade B criterion, it was considered an important and significant omission by members of the BILAG and two nephrologists (Dr Dwomoa Ado and Dr David Jayne) in the consensus group discussion.

Therefore, based on the data presented and the discussions of the consensus group comprising members of the BILAG and two nephrologists experienced in the management of lupus nephritis, the following changes were made to the BILAG-2004 index:

1. Urine protein dipstick result is to be superseded by other methods of urine protein estimation (urine albumin-creatinine ratio, urine protein-creatinine ratio or 24-hours urine protein) where available.
2. The threshold for definition of improvement in proteinuria is changed from 50% to 25%
3. An additional Grade B criterion of urine protein excretion of at least 0.5g/day (or equivalent) that has not improved/decreased by 25%

The new criteria for proteinuria associated with Grades A, B and C of the Renal system is summarised in Table 5-6. This revised BILAG-2004 index was used in the sensitivity to change study.

Table 5-6: New criteria for proteinuria for Grades A, B and C of Renal system for the BILAG-2004 index (changes from the old criteria in italic)

Grade	Criteria for Proteinuria
A	Deteriorating proteinuria (severe) defined as (a) urine protein dipstick increased by ≥ 2 levels *(to be used only if other methods of urine protein estimation are not available)*; **OR** (b) 24-hours urine protein > 1 g that has not decreased (improved) by ≥ *25%*; **OR** (c) urine protein-creatinine ratio > 100 mg/mmol that has not decreased (improved) by ≥ *25%*; **OR** (d) urine albumin-creatinine ratio > 100 mg/mmol that has not decreased (improved) by ≥ *25%*
B	One of the following: (a) Grade A criterion for proteinuria (without fulfilling criteria to score Grade A) (b) urine protein dipstick which has risen by 1 level to at least 2+ *(to be used only if other methods of urine protein estimation are not available)* *(c) 24-hours urine protein of 0.5 - 1g that has not decreased (improved) by ≥ 25%* *(d) urine protein-creatinine ratio of 50 - 100 mg/mmol that has not decreased (improved) by ≥ 25%;* *(e) urine albumin-creatinine ratio of 50 - 100 mg/mmol that has not decreased (improved) by ≥ 25%*
C	Mild/Stable proteinuria defined as (a) urine protein dipstick ≥ 1+ but has not fulfilled criteria for Grades A & B *(to be used only if other methods of urine protein estimation are not available)*; **OR** (b) 24-hours urine protein > 0.25 g but has not fulfilled criteria for Grades A & B ; **OR** (c) urine protein-creatinine ratio > 25 mg/mmol but has not fulfilled criteria for Grades A & B; **OR** (d) urine albumin-creatinine ratio > 25 mg/mmol but has not fulfilled criteria

	for Grades A & B

Chapter 6 Sensitivity to Change Study

6.1 Abstract

Objectives:

1. To determine if the BILAG-2004 index is sensitive to change
2. To compare the sensitivity to change of the BILAG-2004, Classic BILAG and SLEDAI-2000 indices

Methods:

SLE patients were recruited into a prospective multi-centre longitudinal study. At every assessment, data were collected on disease activity (the BILAG-2004, Classic BILAG and SLEDAI-2000 indices), investigations and therapy. Overall BILAG-2004 and Classic BILAG scores were used in the analysis, as determined by the highest score achieved by any of the individual systems. Sensitivity to change was assessed by determining the relationship between change in disease activity and change in therapy between two consecutive visits. Statistical analyses were based on logistic regression.

Results:

A total of 353 SLE patients were recruited and there were 1767 assessments that contributed 1414 observations for analysis. Changes in the BILAG-2004, Classic BILAG and SLEDAI-2000 scores were significantly associated with change in therapy. They were differentially related to change in therapy, with greater change in score having greater predictive power. The BILAG-2004 and Classic BILAG indices had comparable performance when compared. SLEDAI-2000 index retained an independent relationship with change in therapy when compared to the BILAG-2004 and Classic BILAG indices, respectively. Further analysis

showed that changes in SLEDAI-2000 score may be combined with either changes in the BILAG-2004 or Classic BILAG indices in explaining change in therapy.

Conclusions:

1. The BILAG-2004, Classic BILAG and SLEDAI-2000 indices were sensitive to change.
2. There was no superiority of sensitivity to change in any one of the indices when they were compared.
3. Changes in SLEDAI-2000 score might complement changes in the BILAG-2004 or Classic BILAG indices in the analysis of SLE disease activity in a longitudinal study.

6.2 Objectives

1. To determine if the BILAG-2004 index is sensitive to change for the assessment of SLE disease activity.
2. To compare the sensitivity to change of the BILAG-2004, Classic BILAG and SLEDAI-2000 indices

6.3 Introduction

Before the BILAG-2004 index can be used in longitudinal outcome studies of SLE, it needs to be shown that it is sensitive to change. This is important when the index is used to measure transition in clinical states, and in clinical trials, this would affect the number of patients that are required to detect significant effects.

Sensitivity to change or responsiveness is the ability of the index to change with time. There are two forms of responsiveness, namely internal responsiveness and external responsiveness (Husted et al. 2000). Internal responsiveness is defined as the ability of the index to change over a particular time period. External responsiveness refers not just to the ability of the index to change over time but also how the changes in the index relate to the corresponding changes in an external reference. Thus, external responsiveness is a more robust method of assessing sensitivity to change of an index.

A longitudinal study was established to determine if the BILAG-2004 index was sensitive to change. External responsiveness of the index was assessed in this study. As change in therapy was used as the external reference, it is the clinically meaningful change in the index that was being studied. In addition, the sensitivity to change of the Classic BILAG and SLEDAI-2000 indices were also assessed, and comparison of the three indices was made.

6.4 Methods

This was a multi-centre longitudinal study involving 8 centres in the United Kingdom. Recruited SLE patients were followed up prospectively and data were collected for all consecutive visits or encounters (inpatient or outpatient) that the patients had with the physician. Data were collected on disease activity (as assessed using the BILAG-2004, Classic BILAG and SLEDAI-2000 indices), investigations and treatment. This study commenced in March 2005 and was completed in April 2007.

6.5 Analysis

Sensitivity to change analysis was assessed using the method to assess external responsiveness as outlined by Husted et al (Husted et al. 2000). The external reference used in this study was the change in therapy between consecutive visits, which was the difference in treatment after the patient was assessed at the index visit, as compared to the therapy after the previous visit. The definition of change in treatment has been described in Chapter 2. Change in disease activity was defined as the change in the score of the respective indices between the two corresponding consecutive visits.

For the purpose of analysis, overall BILAG-2004 and Classic BILAG scores as determined by the highest score achieved by any of the individual system were used. Scores of D and E were combined together as they both indicate inactivity. Therefore, four categorical overall scores were possible (A, B, C and D).

Maximum-likelihood multinomial logistic regression (with robust variance estimation) was used to assess external responsiveness with change in therapy as the outcome variable and change in disease activity as the explanatory variable. The results were reported in odds ratios (OR) with 95% confidence intervals (CI). The baseline comparator for change in

treatment used in the analysis was ‘no change in therapy’ whereas for change in disease activity, the baseline comparator was ‘no change in activity’ or ‘minimal change in activity’. In addition to ‘no change in activity’, ‘minimal change in activity’ for BILAG-2004 and Classic BILAG indices included change from Grade D to C, as this change was considered minor. For SLEDAI-2000 index, receiver operating characteristic (ROC) curves analysis were used to determine the optimal minimal change in score associated with a change in therapy. Comparison of indices was undertaken within a common regression model, assessing main effects and interactions of the indices as appropriate.

Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were calculated with the same regression methodology, but applied only to appropriate subsets of patients and with only the intercept term estimated. For calculation of sensitivity and its 95% confidence interval, only observations which recorded an increase in treatment were used. Specificity, PPV and NPV were calculated using only patients with no increase in treatment, increase in disease activity and no increase in disease activity respectively.

The relationship between change in SLEDAI-2000 score and change in the BILAG-2004 index score in its association with change in therapy was studied. In this analysis the interaction of the indices were assessed with observations classified into:

1. Increase in SLEDAI-2000 score only (with minimal change in BILAG-2004 score)
2. Increase in BILAG-2004 score only (with minimal change in SLEDAI-2000 score)
3. Increase in both BILAG-2004 and SLEDAI-2000 scores
4. Decrease in SLEDAI-2000 score only (with minimal change in BILAG-2004 score)
5. Decrease in BILAG-2004 score only (with minimal change in SLEDAI-2000 score)
6. Decrease in both BILAG-2004 and SLEDAI-2000 scores

Similar logistic regression to that of sensitivity to change analysis was used, with change in treatment as the outcome variable. The analysis was repeated to define the relationship between change in SLEDAI-2000 score and change in Classic BILAG score.

6.6 Results

A total of 353 SLE patients were recruited and there were 1767 assessments. As 2 consecutive visits or assessments provided 1 observation, there were 1414 observations available for the analysis. The demographics of the patients are summarised in Table 6-1.

Table 6-1: Demographics of patients recruited into the study (n=353).

Patient Characteristics	
Female sex (%)	92.9
Mean age in years (SD)	40.9 (12.9)
Ethnicity (%)	
Caucasian	58.1
Afro-Caribbean	20.1
South Asian	19.0
Oriental	1.4
Others	1.4
Mean disease duration in years (SD)	8.3 (7.9)

There was increase in treatment between consecutive visits in 22.6% of observations while 37.3% had therapy decreased, and in 40.1%, there was no change in treatment. The distribution of the changes in disease activity according to the three indices and change in therapy are summarised in Table 6-2.

Table 6-2: Cross tabulation of changes in disease activity (according to the BILAG-2004, Classic BILAG and SLEDAI-2000 indices) against change in therapy (n=1414).

Change in Disease Activity	Change in Therapy			Total
	Decrease	No Change	Increase	
BILAG-2004				
Decreased (%)	194 (52.3)	146 (39.3)	31 (8.4)	371
No change (%)	280 (36.7)	338 (44.3)	145 (19.0)	763
Increased (%)	53 (18.9)	83 (29.6)	144 (51.4)	280
Classic BILAG				
Decreased (%)	171 (49.7)	139 (40.4)	34 (9.9)	344
No change (%)	313 (38.3)	357 (43.7)	147 (18.0)	873
Increased (%)	43 (17.0)	71 (28.1)	139 (54.9)	197
SLEDAI-2000				
Decrease ≥ 3 (%)	107 (53.0)	68 (33.7)	27 (13.4)	202
Minimal change (%)	391 (36.5)	478 (44.7)	201 (18.8)	1070
Increase ≥ 3 (%)	29 (20.4)	21 (14.8)	92 (64.8)	142

There was no significant effect of duration of follow-up ($p>0.4$) but there was a higher likelihood of decrease in therapy with greater number of visits (OR 1.10, 95% CI 1.04, 1.16).

6.6.1 Sensitivity to Change of the BILAG-2004 Index

Increase in overall BILAG-2004 score was significantly associated with increase in therapy. On the other hand, decrease in overall BILAG-2004 score was significantly associated with decrease in therapy and inversely associated with increase in therapy (Table 6-3).

Table 6-3: Sensitivity to change analysis of the BILAG-2004 index determining the association of change in overall BILAG-2004 score with change in therapy (n=1414).

Change in Overall BILAG-2004 Score	Number of Observations	Increase in Therapy Odds Ratio (95% CI)	Decrease in Therapy Odds Ratio (95% CI)
No change in activity	763	1.0	1.0
Increase in activity	280	4.0 (2.9, 5.7)	0.8 (0.5, 1.1)
Decrease in activity	371	0.5 (0.3, 0.8)	1.6 (1.2, 2.1)

As Grade C was considered as mild activity, a change of grade from D/E (inactivity) to C was considered minor. When this change of score from D/E to C was excluded from the definition of increase in activity (considered minimal change in activity), increase in overall BILAG-2004 scores had a much greater predictive power of increase in therapy (Table 6-4).

Table 6-4: Sensitivity to change analysis of the BILAG-2004 index after excluding change of Grade D/E to C from the definition of increase in activity (n=1414).

Change in Overall BILAG-2004 Score	Number of Observation	Increase in Therapy Odds Ratio (95% CI)	Decrease in Therapy Odds Ratio (95% CI)
Minimal change in activity	848	1.0	1.0
Increase in activity	195	10.7 (6.8, 16.6)	1.1 (0.7, 1.8)
Decrease in activity	371	0.5 (0.3, 0.8)	1.7 (1.3, 2.2)

Further analysis on the subgroups of changes in score revealed that increase in the score to Grade A was more likely than increase in the score to Grade B to be associated with increase in therapy, whereas decrease in the score to Grade C or D/E was much less likely to have increase in therapy than decrease in the score to Grade B (Table 6-5). Therefore, changes in overall BILAG-2004 score were differentially related to change in therapy, with greater change in score having greater predictive power.

Table 6-5: Analysis on the subgroups of changes in overall BILAG-2004 score and its association with change in therapy (n=1414).

Change in Overall BILAG-2004 Score	Number of Observations	Increase in Therapy Odds Ratio (95% CI)	Decrease in Therapy Odds Ratio (95% CI)
Minimal change in activity	848	1.0	1.0
Increase in activity to Grade A	48	16.4 (6.0, 45.0)	0.6 (0.1, 2.8)
Increase in activity to Grade B	147	9.3 (5.7, 15.3)	1.2 (0.7, 2.1)
Decrease in activity to Grade B	41	0.7 (0.2, 2.0)	1.9 (0.9, 3.7)
Decrease in activity to Grade C or D	330	0.5 (0.3, 0.8)	1.7 (1.3, 2.2)

6.6.2 Sensitivity to Change of Classic BILAG Index

Increase in overall Classic BILAG score was significantly associated with increase in therapy while decrease in overall score was significantly associated with decrease in therapy and inversely associated with increase in therapy (Table 6-6). Similar to the BILAG-2004 index, exclusion of change in the score of D to C from the definition of increase in activity increased the predictive power of overall Classic BILAG score on increase in therapy (Table 6-7). Changes in overall Classic BILAG score were also differentially related to change in therapy, with greater change in score having greater predictive power (Table 6-8).

Table 6-6: Sensitivity to change analysis of Classic BILAG index determining the association of change in overall Classic BILAG score with change in therapy (n=1414).

Change in Overall Classic BILAG Score	Number of Observations	Increase in Therapy Odds Ratio (95% CI)	Decrease in Therapy Odds Ratio (95% CI)
No change in activity	817	1.0	1.0
Increase in activity	253	4.8 (3.4, 6.7)	0.7 (0.5, 1.0)
Decrease in activity	344	0.6 (0.4, 0.9)	1.4 (1.1, 1.8)

Table 6-7: Sensitivity to change analysis of Classic BILAG index after excluding change of Grade D/E to C from the definition of increase in activity (n=1414).

Change in Overall Classic BILAG Score	Number of Observation	Increase in Therapy Odds Ratio (95% CI)	Decrease in Therapy Odds Ratio (95% CI)
Minimal change in activity	873	1.0	1.0
Increase in activity	197	9.5 (6.2, 14.4)	0.9 (0.6, 1.5)
Decrease in activity	344	0.6 (0.4, 0.9)	1.5 (1.1, 1.9)

Table 6-8: Analysis on the subgroups of changes in overall Classic BILAG score and its association with change in therapy (n=1414).

Change in Overall Classic BILAG Score	Number of Observations	Increase in Therapy Odds Ratio (95% CI)	Decrease in Therapy Odds Ratio (95% CI)
Minimal change in activity	873	1.0	1.0
Increase in activity to Grade A	52	31.3 (9.5, 102.9)	1.5 (0.3, 6.9)
Increase in activity to Grade B	145	6.6 (4.2, 10.4)	0.9 (0.6, 1.6)
Decrease in activity to Grade B	40	0.7 (0.3, 1.9)	1.5 (0.8, 3.0)
Decrease in activity to Grade C or D	304	0.7 (0.4, 1.0)	1.5 (1.2, 2.0)

6.6.3 Sensitivity to Change of SLEDAI-2000

Increase in SLEDAI-2000 score was significantly associated with increase in therapy (OR 1.2, 95% CI 1.2 – 1.3) and inversely associated with treatment reduction (OR 0.9, 95% CI 0.9 – 1.0). When the SLEDAI-2000 score of the previous visit was also included in the regression model, both the change in SLEDAI-2000 score and the previous visit SLEDAI-2000 score were significantly associated with increase in therapy (Table 6-9). Therefore, the model with just change in SLEDAI-2000 score was insufficient to explain change in therapy (especially treatment increase).

Table 6-9: Association between change in therapy with change in SLEDAI-2000 score and SLEDAI-2000 score of previous visit.

	Increase in Therapy Odds Ratio* (95% CI)	**Decrease in Therapy Odds Ratio* (95% CI)**
Change in SLEDAI-2000 score	1.4 (1.3, 1.5)	0.9 (0.9, 1.0)
Previous visit SLEDAI-2000 score	1.2 (1.2, 1.3)	1.0 (0.9, 1.1)

* per unit change in SLEDAI-2000 score

ROC curves analysis with SLEDAI-2000 index as a predictor of increase in therapy, to determine the optimal increase in score, is summarised in Table 6-10. When all the parameters were considered, the optimal minimal change in SLEDAI-2000 score associated with increase in therapy was 1 or 2. However, for the best performance in predicting increase in therapy, minimal change in SLEDAI-2000 score of 3 or 4 was preferable as both had better positive predictive values (as compared to minimal change of score of 2) although there was a compromise in the sensitivity.

ROC curves analysis was also performed to determine the optimal decrease in score that was associated with no increase in therapy (combination of decrease in therapy and no change in therapy) as shown in Table 6-11. From the analysis, the optimal decrease in SLEDAI-2000 score associated with no increase in treatment was 1 or 2. The results were similar when the analysis with decrease in therapy was used instead of no increase in treatment (data not shown).

Table 6-10: Receiver operating characteristic curves analysis of SLEDAI-2000 index to determine the optimal increase in score associated with increase in therapy.

Increase in SLEDAI-2000 Score	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
≥ 1	47.5% (41.7, 53.3)	82.4% (80.0, 84.6)	44.2% (39.0, 49.5)	84.3% (81.7, 86.6)
≥ 2	45.0% (39.5, 50.7)	84.4% (82.2, 86.3)	45.7% (40.4, 51.2)	84.0% (81.5, 86.2)
≥ 3	28.8% (23.8, 34.3)	95.4% (94.0, 96.5)	64.8% (56.8, 72.0)	82.1% (79.6, 84.3)
≥ 4	26.3% (21.5, 31.7)	95.9% (94.4, 96.9)	65.1% (56.5, 72.8)	81.6% (79.1, 83.9)
≥ 5	14.7% (10.6, 19.9)	98.5% (97.4, 99.1)	73.4% (60.8, 83.1)	79.8% (77.2, 82.1)
≥ 6	13.8% (9.8, 19.0)	98.8% (97.9, 99.3)	77.2% (64.5, 86.3)	79.7% (77.1, 82.0)
≥ 7	9.1% (6.1, 13.4)	99.5% (98.7, 99.8)	85.3% (67.7, 94.1)	78.9% (76.4, 81.2)

Table 6-11: Receiver operating characteristic curves analysis of SLEDAI-2000 index to determine the optimal decrease in score associated with no increase in therapy.

Decrease in SLEDAI-2000 Score	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
≥ 1	31.7% (28.9, 34.7)	80.6% (76.3, 84.3)	84.8% (81.1, 87.9)	25.7% (22.7, 28.9)
≥ 2	30.3% (27.6, 33.3)	81.9% (77.6, 85.5)	85.1% (81.3, 88.3)	25.6% (22.7, 28.7)
≥ 3	16.0% (13.8, 18.5)	91.6% (88.1, 94.1)	86.6% (81.4, 90.6)	24.2% (21.5, 27.0)
≥ 4	14.9% (12.7, 17.4)	93.1% (89.9, 95.4)	88.1% (82.8, 91.9)	24.2% (21.6, 27.1)
≥ 5	7.2% (5.6, 9.2)	96.6% (94.1, 98.0)	87.8% (80.1, 92.8)	23.3% (20.9, 26.0)
≥ 6	7.0% (5.5, 9.0)	97.2% (94.8, 98.5)	89.5% (81.8, 94.2)	23.4% (20.9, 26.1)
≥ 7	4.7% (3.4, 6.4)	98.1% (96.0, 99.1)	89.5% (79.7, 94.8)	23.1% (20.7, 25.8)

When change in SLEDAI-2000 score was defined as minimal increase of 2 to indicate worsening and minimal decrease of 2 to indicate improvement (Model A), increase in SLEDAI-2000 score was significantly associated with increase in therapy and decrease in score was associated with treatment reduction (Table 6-12). The results were very similar if the minimal increase in score of 1 or minimal decrease in score of 1 were used instead (data not shown).

With a different model of change in SLEDAI-2000 score using minimal increase of 3 (Model B, minimal increase of 3 and minimal decrease of 1), increase in SLEDAI-2000 had a much stronger association with increase in therapy than Model A (Table 6-13). The results

were similar with minimal increase of 4 or minimal decrease of 2 (data not shown). This is consistent with the recommendation of using minimal increase in score of 3 or 4 instead of 1 or 2.

Table 6-12: Sensitivity to change analysis of SLEDAI-2000 index with Model A of change in score (minimal increase of 2 and minimal decrease of 2) (n=1414).

Change in SLEDAI-2000 Score	Number of Observation	Increase in Therapy Odds Ratio (95% CI)	Decrease in Therapy Odds Ratio (95% CI)
Minimal change	709	1.0	1.0
Increase ≥ 2	315	4.8 (3.4, 6.9)	1.4 (0.9, 1.9)
Decrease ≥ 2	390	1.3 (0.9, 1.9)	2.2 (1.7, 2.9)

Table 6-13: Sensitivity to change analysis of SLEDAI-2000 index with Model B of change in score (minimal increase of 3 and minimal decrease of 1) (n=1414).

Change in SLEDAI-2000 Score	Number of Observation	Increase in Therapy Odds Ratio (95% CI)	Decrease in Therapy Odds Ratio (95% CI)
Minimal change	863	1.0	1.0
Increase ≥ 3	142	10.6 (6.2, 18.1)	1.9 (1.0, 3.5)
Decrease ≥ 1	409	1.1 (0.7, 1.5)	2.0 (1.5, 2.6)

Further exploratory analysis with alternative models of change in SLEDAI-2000 score was performed. Firstly, the magnitude of change was classified into ordinal categories based broadly on the minimal increase in score of 3 (for worsening) and minimal decrease in score of 1 (for improvement) (Model C):

1. Minimal change – no change in score or increase in score of less than 3
2. Increase in score of up to twice the minimal increase in score (3 to 6)
3. Increase in score of more than twice the minimal increase in score (more than 6)
4. Decrease in score of up to twice the minimal decrease in score (1 to 2)

5. Decrease in score of more than twice the minimal decrease in score (more than 2)

With this model (Model C), changes in SLEDAI-2000 score were differentially related to change in therapy with greater change in score having greater predictive power of change in therapy (Table 6-14).

Table 6-14: Analysis on the subgroups of changes in SLEDAI-2000 score based on magnitude of change (Model C) and its association with change in therapy (n=1414).

Change in SLEDAI-2000 Score	Number of Observation	Increase in Therapy Odds Ratio (95% CI)	Decrease in Therapy Odds Ratio (95% CI)
Minimal change	863	1.0	1.0
Increase by more than 6	34	35.2 (8.5, 145.5)	2.1 (0.5, 8.3)
Increase by 3 to 6	108	8.0 (4.6, 14.1)	1.9 (1.0, 3.6)
Decrease by 1 to 2	207	1.1 (0.7, 1.7)	1.8 (1.3, 2.5)
Decrease by more than 2	202	1.0 (0.6, 1.6)	2.2 (1.5, 3.0)

Another way to assess change in score was to incorporate SLEDAI-2000 score of the previous visit (which had been shown to be important in explaining change in therapy) whereby SLEDAI-2000 score is classified into ordinal categories in similar fashion to the BILAG-2004 index. The classification of the ordinal categories is based on the cut-off score for active disease of 3 (as shown in the cross sectional study): score of less than 3, score of 3 to 7 and score of more than 7. Similar to the previous analysis, changes in SLEDAI-2000 score using this categorical model (Model D) were differentially related to changes in therapy (Table 6-15).

Table 6-15: Analysis on the subgroups of categorical changes in SLEDAI-2000 score (Model D) and its association with change in therapy (n=1414).

Change in SLEDAI-2000 Score	Number of Observation	Increase in Therapy Odds Ratio (95% CI)	Decrease in Therapy Odds Ratio (95% CI)
Minimal change	999	1.0	1.0
Increase to score of > 7	68	15.5 (6.6, 36.5)	1.2 (0.4, 3.5)
Increase to score of 3 to 7	120	4.0 (2.5, 6.4)	1.2 (0.7, 2.0)
Decrease to score of 3 to 7	61	1.8 (0.8, 3.7)	1.8 (1.0, 3.2)
Decrease to score of < 3	166	0.5 (0.2, 1.0)	1.9 (1.3, 2.7)

Thus far, several models of change in SLEDAI-2000 score had been examined. Comparison of these models of change in SLEDAI-2000 score was made using area under the ROC curve with increase in therapy as the outcome variable (Table 6-16). The model with change in SLEDAI-2000 score as a continuous variable and the SLEDAI-2000 score of the previous visit included, had the best performance in explaining increase in therapy while all other models performed similarly.

Table 6-16: Area under the receiver operating characteristic curve for the various models of change in SLEDAI-2000 score in explaining increase in therapy.

Models of change in SLEDAI-2000 Score	Area under ROC curve
Change in score (continuous variable)	0.67
Change in score (continuous variable) and previous visit SLEDAI-2000 score	0.76
Model A (minimal increase of 2 and minimal decrease of 2)	0.65
Model B (minimal increase of 3 and minimal decrease of 1)	0.64
Model C (based on magnitude of change)	0.65
Model D (categorical changes)	0.67

6.6.4 Comparison of Indices

(Tables 6-17 to 6-30 at the end of section)

When changes in the score for both the BILAG-2004 and Classic BILAG indices were included in a common regression model, increase in the score for both indices were significantly associated with an increase in therapy (Table 6-17). However, only decrease in BILAG-2004 score was significantly associated with a decrease in therapy, which was not the case for decrease in Classic BILAG score. Further comparison of these two indices, by assessing the ability of increase in the score of the respective indices in predicting increase in therapy, showed that they had similar sensitivity, specificity, PPV and NPV (Table 6-18).

Comparison was made between changes in BILAG-2004 score and the various models of change in SLEDAI-2000 score (Tables 6-19 to 6-24). When changes in the score of both the BILAG-2004 and SLEDAI-2000 indices were included within a common regression model, both of the indices remained significantly associated with change in therapy, regardless of the model of change in SLEDAI-2000 score used in the analysis. Therefore, both the BILAG-2004 and SLEDAI-2000 indices retained independent relationships with change in therapy when both were included in the regression model. Similarly, changes in the score of both Classic BILAG and SLEDAI-2000 indices had independent relationships with change in therapy when these two indices were compared (Tables 6-25 to 6-30).

Therefore, the comparison of the indices did not show that any of the three indices was clearly superior with regards to sensitivity to change. In this analysis, the BILAG-2004 index had comparable performance to Classic BILAG index. There was some non-overlapping relationship with change in therapy when SLEDAI-2000 index was compared with the BILAG-2004 and Classic BILAG indices, respectively.

Table 6-17: Comparison of the ability of change in BILAG-2004 score and change in Classic BILAG score in predicting change in therapy (n=1414).

Changes in Score	Increase in Therapy Odds Ratio (95% CI)	Decrease in Therapy Odds Ratio (95% CI)
Overall BILAG-2004 score		
Increase (excluding D to C)	4.7 (2.8, 8.0)	1.3 (0.7, 2.4)
Decrease	0.7 (0.4, 1.1)	1.7 (1.2, 2.4)
Overall Classic BILAG score		
Increase (excluding D to C)	3.2 (1.9, 5.3)	0.9 (0.5, 1.7)
Decrease	0.9 (0.6, 1.5)	1.1 (0.8, 1.5)

Table 6-18: Sensitivity, specificity, positive predictive value and negative predictive value for the model of increase in activity, according to the BILAG-2004 and Classic BILAG indices, in predicting increase in therapy.

Indices	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)
BILAG-2004	42.2 (37.0, 47.6)	94.5 (92.9, 95.8)	69.2 (61.9, 75.7)	84.8 (82.4, 87.0)
Classic BILAG	41.6 (36.6, 46.7)	94.1 (92.6, 95.4)	67.5 (60.8, 73.6)	84.6 (82.2, 86.8)

Table 6-19: Comparison of the ability of change in BILAG-2004 score and change in SLEDAI-2000 score (continuous variable) in predicting change in therapy (n=1414).

Changes in Score	Increase in Therapy Odds Ratio (95% CI)	Decrease in Therapy Odds Ratio (95% CI)
Overall BILAG-2004 score		
Increase to Grade A	13.0 (4.7, 36.0)	0.7 (0.2, 3.2)
Increase to Grade B	7.5 (4.4, 12.6)	1.4 (0.8, 2.4)
Decrease to Grade B	0.8 (0.3, 2.6)	1.6 (0.8, 3.1)
Decrease to Grade C/D	0.6 (0.4, 1.0)	1.5 (1.2, 2.0)
Change in SLEDAI-2000 score (continuous variable)	1.1 (1.0, 1.2)	0.9 (0.9, 1.0)

Table 6-20: Comparison of the ability of change in BILAG-2004 score and change in SLEDAI-2000 score (continuous variable) with previous visit SLEDAI-2000 score, in predicting change in therapy (n=1414).

Changes in Score	Increase in Therapy Odds Ratio (95% CI)	Decrease in Therapy Odds Ratio (95% CI)
Overall BILAG-2004 score		
Increase to Grade A	8.4 (4.0, 11.5)	0.7 (0.2, 3.1)
Increase to Grade B	6.7 (4.0, 11.5)	1.4 (0.8, 2.3)
Decrease to Grade B	0.3 (0.1, 1.2)	1.5 (0.8, 3.0)
Decrease to Grade C/D	0.6 (0.3, 0.9)	1.5 (1.1, 2.0)
Change in SLEDAI-2000 score (continuous variable)	1.2 (1.1, 1.3)	0.9 (0.9, 1.0)
Previous visit SLEDAI-2000 score	1.2 (1.1, 1.3)	1.0 (0.9, 1.1)

Table 6-21: Comparison of the ability of change in BILAG-2004 score and Model A of change in SLEDAI-2000 score in predicting change in therapy (n=1414).

Changes in Score	Increase in Therapy Odds Ratio (95% CI)	Decrease in Therapy Odds Ratio (95% CI)
Overall BILAG-2004 score		
Increase to Grade A	12.8 (4.5, 36.0)	0.6 (0.1, 2.6)
Increase to Grade B	6.7 (4.0, 11.1)	1.1 (0.7, 2.0)
Decrease to Grade B	0.6 (0.2, 1.9)	1.5 (0.8, 2.8)
Decrease to Grade C/D	0.5 (0.3, 0.9)	1.4 (1.1, 1.9)
SLEDAI-2000 score		
Increase (≥ 2)	2.4 (1.6, 3.5)	1.4 (1.0, 2.0)
Decrease (≥ 2)	1.6 (1.1, 2.5)	2.0 (1.5, 2.6)

Table 6-22: Comparison of the ability of change in BILAG-2004 score and Model B of change in SLEDAI-2000 score in predicting change in therapy (n=1414).

Changes in Score	Increase in Therapy Odds Ratio (95% CI)	Decrease in Therapy Odds Ratio (95% CI)
Overall BILAG-2004 score		
Increase to Grade A	11.4 (4.0, 32.6)	0.6 (0.1, 2.5)
Increase to Grade B	6.1 (3.6, 10.3)	1.1 (0.6, 1.9)
Decrease to Grade B	0.6 (0.2, 1.9)	1.5 (0.8, 2.9)
Decrease to Grade C/D	0.5 (0.3, 0.8)	1.5 (1.1, 1.9)
SLEDAI-2000 score		
Increase (≥ 3)	4.4 (2.4, 8.0)	2.0 (1.1, 3.7)
Decrease (≥ 1)	1.5 (1.0, 2.2)	1.7 (1.3, 2.3)

Table 6-23: Comparison of the ability of change in BILAG-2004 score and Model C of change in SLEDAI-2000 score in predicting change in therapy (n=1414).

Changes in Score	Increase in Therapy Odds Ratio (95% CI)	Decrease in Therapy Odds Ratio (95% CI)
Overall BILAG-2004 score		
Increase to Grade A	9.8 (3.6, 26.9)	0.3 (0.1, 1.6)
Increase to Grade B	5.2 (3.1, 8.5)	0.8 (0.4, 1.4)
Decrease to Grade B	0.7 (0.2, 1.8)	1.1 (0.6, 2.1)
Decrease to Grade C/D	0.4 (0.3, 0.8)	1.4 (1.1, 1.9)
SLEDAI-2000 score		
Increase by more than 6	12.3 (2.5, 59.8)	2.5 (0.6, 9.5)
Increase by 3 to 6	4.0 (2.1, 7.3)	2.1 (1.1, 4.1)
Decrease by 1 to 2	1.5 (0.9, 2.3)	1.7 (1.2, 2.3)
Decrease by more than 2	1.4 (0.8, 2.5)	1.9 (1.3, 2.7)

Table 6-24: Comparison of the ability of change in BILAG-2004 score and Model D of change in SLEDAI-2000 score in predicting change in therapy (n=1414).

Changes in Score	Increase in Therapy Odds Ratio (95% CI)	Decrease in Therapy Odds Ratio (95% CI)
Overall BILAG-2004 score		
Increase to Grade A	10.8 (3.8, 30.4)	0.6 (0.1, 2.7)
Increase to Grade B	6.1 (3.6, 10.3)	1.2 (0.7, 2.0)
Decrease to Grade B	0.6 (0.2, 2.0)	1.7 (0.9, 3.3)
Decrease to Grade C/D	0.6 (0.3, 0.9)	1.5 (1.1, 2.0)
SLEDAI-2000 score		
Increase to score of > 7	4.8 (1.7, 13.1)	1.3 (0.5, 3.7)
Increase to score of 3 to 7	2.2 (1.3, 3.6)	1.2 (0.7, 2.1)
Decrease to score of 3 to 7	2.3 (1.1, 5.1)	1.5 (0.9, 2.7)
Decrease to score of < 3	0.6 (0.3, 1.3)	1.6 (1.2, 2.4)

Table 6-25: Comparison of the ability of change in Classic BILAG score and change in SLEDAI-2000 score (continuous variable) in predicting change in therapy (n=1414).

Changes in Score	Increase in Therapy Odds Ratio (95% CI)	Decrease in Therapy Odds Ratio (95% CI)
Overall Classic BILAG score		
Increase to Grade A	22.3 (7.7, 64.6)	1.0 (0.2, 4.4)
Increase to Grade B	5.5 (3.5, 8.8)	1.0 (0.6, 1.8)
Decrease to Grade B	1.3 (0.5, 3.7)	1.2 (0.6, 2.4)
Decrease to Grade C/D	0.7 (0.5, 1.2)	1.3 (1.0, 1.7)
Change in SLEDAI-2000 score (continuous variable)	1.1 (1.1, 1.2)	0.9 (0.9, 1.0)

Table 6-26: Comparison of the ability of change in Classic BILAG score and change in SLEDAI-2000 score (continuous variable) with previous visit SLEDAI-2000 score, in predicting changes in therapy (n=1414).

Changes in Score	Increase in Therapy Odds Ratio (95% CI)	Decrease in Therapy Odds Ratio (95% CI)
Overall Classic BILAG score		
Increase to Grade A	16.1 (5.3, 48.7)	1.0 (0.2, 4.5)
Increase to Grade B	5.6 (3.4, 9.1)	1.0 (0.6, 1.8)
Decrease to Grade B	0.5 (0.2, 1.5)	1.2 (0.6, 2.3)
Decrease to Grade C/D	0.7 (0.4, 1.1)	1.3 (1.0, 1.7)
Change in SLEDAI-2000 score (continuous variable)	1.2 (1.2, 1.3)	0.9 (0.9, 1.0)
Previous visit SLEDAI-2000 score	1.2 (1.2, 1.3)	1.0 (1.0, 1.1)

Table 6-27: Comparison of the ability of change in Classic BILAG score and Model A of change in SLEDAI-2000 score in predicting changes in therapy (n=1414).

Changes in Score	Increase in Therapy Odds Ratio (95% CI)	Decrease in Therapy Odds Ratio (95% CI)
Overall Classic BILAG score		
Increase to Grade A	24.0 (8.1, 71.4)	0.9 (0.2, 4.1)
Increase to Grade B	5.1 (3.2, 8.1)	0.9 (0.5, 1.6)
Decrease to Grade B	0.8 (0.3, 2.3)	1.2 (0.6, 2.3)
Decrease to Grade C/D	0.6 (0.4, 1.0)	1.2 (0.9, 1.6)
SLEDAI-2000 score		
Increase (≥ 2)	2.7 (1.9, 4.0)	1.3 (0.9, 1.8)
Decrease (≥ 2)	1.6 (1.1, 2.5)	2.0 (1.5, 2.7)

Table 6-28: Comparison of the ability of change in Classic BILAG score and Model B of change in SLEDAI-2000 score in predicting change in therapy (n=1414).

Changes in Score	Increase in Therapy Odds Ratio (95% CI)	Decrease in Therapy Odds Ratio (95% CI)
Overall Classic BILAG score		
Increase to Grade A	21.8 (7.6, 62.6)	0.9 (0.2, 3.9)
Increase to Grade B	4.8 (3.0, 7.8)	0.9 (0.5, 1.5)
Decrease to Grade B	0.8 (0.3, 2.3)	1.2 (0.6, 2.3)
Decrease to Grade C/D	0.6 (0.4, 1.0)	1.3 (1.0, 1.7)
SLEDAI-2000 score		
Increase (≥ 3)	5.8 (3.3, 10.2)	2.0 (1.1, 3.7)
Decrease (≥ 1)	1.4 (0.9, 2.1)	1.8 (1.4, 2.4)

Table 6-29: Comparison of the ability of change in Classic BILAG score and Model C of change in SLEDAI-2000 score in predicting change in therapy (n=1414).

Changes in Score	Increase in Therapy Odds Ratio (95% CI)	Decrease in Therapy Odds Ratio (95% CI)
Overall Classic BILAG score		
Increase to Grade A	21.2 (7.4, 60.9)	0.9 (0.2, 4.0)
Increase to Grade B	5.0 (3.1, 8.1)	0.9 (0.5, 1.5)
Decrease to Grade B	0.8 (0.3, 2.4)	1.2 (0.6, 2.3)
Decrease to Grade C/D	0.6 (0.4, 1.0)	1.2 (0.9, 1.6)
SLEDAI-2000 score		
Increase by more than 6	18.2 (4.0, 82.9)	2.2 (0.5, 8.8)
Increase by 3 to 6	4.5 (2.5, 8.2)	2.0 (1.0, 3.8)
Decrease by 1 to 2	1.3 (0.8, 2.1)	1.7 (1.2, 2.4)
Decrease by more than 2	1.5 (0.8, 2.6)	2.0 (1.4, 2.8)

Table 6-30: Comparison of the ability of change in Classic BILAG score and Model D of change in SLEDAI-2000 score in predicting change in therapy (n=1414).

Changes in Score	Increase in Therapy Odds Ratio (95% CI)	Decrease in Therapy Odds Ratio (95% CI)
Overall Classic BILAG score		
Increase to Grade A	21.5 (7.5, 61.9)	0.9 (0.2, 4.2)
Increase to Grade B	5.1 (3.1, 8.3)	0.9 (0.5, 1.5)
Decrease to Grade B	0.9 (0.3, 2.5)	1.4 (0.7, 2.7)
Decrease to Grade C/D	0.7 (0.4, 1.1)	1.3 (1.0, 1.7)
SLEDAI-2000 score		
Increase to score of > 7	8.7 (3.4, 22.1)	1.3 (0.5, 3.8)
Increase to score of 3 to 7	2.2 (1.3, 3.7)	1.3 (0.7, 2.2)
Decrease to score of 3 to 7	2.2 (1.0, 4.7)	1.6 (0.9, 2.9)
Decrease to score of < 3	0.6 (0.3, 1.4)	1.7 (1.2, 2.5)

6.6.5 Relationship between SLEDAI-2000 with the BILAG-2004 and Classic BILAG indices

(Tables 6-31 to 6-38 at the end of section)

In view of the non-overlapping relationship between change in SLEDAI-2000 score and change in BILAG-2004 score with change in therapy, exploratory analysis of this relationship was performed. Several thresholds for definition of change in SLEDAI-2000 score were used in the analysis for comparison (Tables 6-31 to 6-34).

For increase in therapy, a consistent increase in both BILAG-2004 and SLEDAI-2000 scores had the strongest effect, with an increase in BILAG-2000 score only and an increase in SLEDAI-2000 score only having lesser effects which could not be differentiated statistically. This applied across the four thresholds of increase in SLEDAI-2000 scores. Increase in SLEDAI-2000 score of at least 3 performed the best when compared to the other three thresholds as it had the strongest association with increase in therapy.

As for decrease in therapy, the strongest association occurred when there was decrease in the scores of both indices. Decrease in SLEDAI-2000 score on its own had a similar effect to decrease in BILAG-2004 score only. The results were similar for the four thresholds of decrease in SLEDAI-2000 score. Therefore, when the BILAG-2004 index is to be used in combination with SLEDAI-2000 index, the optimal performance was with the use of minimal increase of 3 and minimal decrease of 1 or 2 in the definition of change in SLEDAI-2000 score.

Similar analysis was performed to define the relationship between Classic BILAG index and SLEDAI-2000 (Tables 6-35 to 6-38). The results were similar to the BILAG-2004 index in which the best performance for the combination of Classic BILAG and SLEDAI-2000 indices was when minimal increase of 3 and minimal decrease of 1 or 2 was used in the definition of change in SLEDAI-2000 score.

Table 6-31: Relationship between change in SLEDAI-2000 score (minimal change of 1) and change in BILAG-2004 score in explaining change in therapy (n=1414).

Change in Score	Number of Observation	Increase in Therapy Odds Ratio (95% CI)	Decrease in Therapy Odds Ratio (95% CI)
Increase (≥ 1) in SLEDAI-2000 score only	207	2.3 (1.4, 3.6)	1.3 (0.9, 1.9)
Increase in BILAG-2004 score only (exclude D to C)	58	8.6 (4.1, 18.2)	1.3 (0.7, 2.6)
Increase in both BILAG-2004 and SLEDAI-2000 scores	137	17.9 (10.3, 31.1)	1.2 (0.6, 2.4)
Decrease (≥ 1) in SLEDAI-2000 score only	207	1.9 (1.2, 3.0)	1.7 (1.2, 2.5)
Decrease in BILAG-2004 score only	169	0.7 (0.4, 1.3)	1.4 (1.0, 2.0)
Decrease in both SLEDAI-2000 and BILAG-2004 scores	202	0.7 (0.3, 1.3)	2.6 (1.8, 3.8)

Table 6-32: Relationship between change in SLEDAI-2000 score (minimal change of 2) and change in BILAG-2004 score in explaining change in therapy (n=1414).

Change in Score	Number of Observation	Increase in Therapy Odds Ratio (95% CI)	Decrease in Therapy Odds Ratio (95% CI)
Increase (≥ 2) in SLEDAI-2000 score only	183	2.3 (1.4, 3.7)	1.4 (1.0, 2.1)
Increase in BILAG-2004 score only (exclude D to C)	63	7.6 (3.8, 15.2)	1.3 (0.6, 2.4)
Increase in both BILAG-2004 and SLEDAI-2000 scores	132	18.8 (10.5, 33.4)	1.3 (0.6, 2.8)
Decrease (≥ 2) in SLEDAI-2000 score only	195	1.9 (1.2, 3.0)	1.9 (1.3, 2.7)
Decrease in BILAG-2004 score only	176	0.7 (0.4, 1.2)	1.4 (1.0, 1.9)
Decrease in both SLEDAI-2000 and BILAG-2004 scores	195	0.6 (0.3, 1.3)	2.9 (2.0, 4.2)

Table 6-33: Relationship between change in SLEDAI-2000 score (minimal change of 3) and change in BILAG-2004 score in explaining change in therapy (n=1414).

Change in Score	Number of Observation	Increase in Therapy Odds Ratio (95% CI)	Decrease in Therapy Odds Ratio (95% CI)
Increase (≥ 3) in SLEDAI-2000 score only	52	6.9 (3.3, 14.7)	2.0 (0.9, 4.4)
Increase in BILAG-2004 score only (exclude D to C)	105	9.0 (5.3, 15.3)	1.0 (0.5, 1.8)
Increase in both BILAG-2004 and SLEDAI-2000 scores	90	20.8 (9.6, 45.0)	1.7 (0.6, 4.3)
Decrease (≥ 3) in SLEDAI-2000 score only	77	2.1 (1.1, 3.9)	1.9 (1.1, 3.3)
Decrease in BILAG-2004 score only	246	0.7 (0.4, 1.1)	1.6 (1.2, 2.2)
Decrease in both SLEDAI-2000 and BILAG-2004 scores	125	0.5 (0.2, 1.1)	2.3 (1.5, 3.5)

Table 6-34: Relationship between change in SLEDAI-2000 score (minimal change of 4) and change in BILAG-2004 score in explaining change in therapy (n=1414).

Change in Score	Number of Observation	Increase in Therapy Odds Ratio (95% CI)	Decrease in Therapy Odds Ratio (95% CI)
Increase (≥ 4) in SLEDAI-2000 score only	44	5.5 (2.6, 11.6)	1.5 (0.7, 3.5)
Increase in BILAG-2004 score only (exclude D to C)	110	8.8 (5.3, 14.8)	0.9 (0.5, 1.7)
Increase in both BILAG-2004 and SLEDAI-2000 scores	85	20.9 (9.3, 47.3)	1.8 (0.7, 4.8)
Decrease (≥ 4) in SLEDAI-2000 score only	70	1.9 (1.0, 3.6)	1.8 (1.0, 3.3)
Decrease in BILAG-2004 score only	256	0.7 (0.5, 1.2)	1.6 (1.2, 2.2)
Decrease in both SLEDAI-2000 and BILAG-2004 scores	115	0.3 (0.1, 0.8)	2.2 (1.5, 3.4)

Table 6-35: Relationship between change in SLEDAI-2000 score (minimal change of 1) and change in Classic BILAG score in explaining change in therapy (n=1414).

Change in Score	Number of Observation	Increase in Therapy Odds Ratio (95% CI)	Decrease in Therapy Odds Ratio (95% CI)
Increase (≥ 1) in SLEDAI-2000 score only	218	2.9 (1.9, 4.6)	1.2 (0.8, 1.8)
Increase in Classic BILAG score only (exclude D to C)	71	8.9 (4.7, 16.6)	1.0 (0.5, 2.1)
Increase in both Classic BILAG and SLEDAI-2000 scores	126	17.9 (10.3, 31.1)	1.1 (0.5, 2.2)
Decrease (≥ 1) in SLEDAI-2000 score only	221	1.9 (1.2, 3.1)	1.8 (1.3, 2.6)
Decrease in Classic BILAG score only	156	0.9 (0.5, 1.5)	1.2 (0.8, 1.8)
Decrease in both SLEDAI-2000 and Classic BILAG scores	188	0.8 (0.4, 1.5)	2.3 (1.6, 3.3)

Table 6-36: Relationship between change in SLEDAI-2000 score (minimal change of 2) and change in Classic BILAG score in explaining change in therapy (n=1414).

Change in Score	Number of Observation	Increase in Therapy Odds Ratio (95% CI)	Decrease in Therapy Odds Ratio (95% CI)
Increase (≥ 2) in SLEDAI-2000 score only	194	2.8 (1.8, 4.4)	1.4 (0.9, 2.0)
Increase in Classic BILAG score only (exclude D to C)	76	7.0 (3.9, 12.5)	0.8 (0.4, 1.6)
Increase in both Classic BILAG and SLEDAI-2000 scores	121	21.0 (11.8, 37.4)	1.4 (0.7, 2.8)
Decrease (≥ 2) in SLEDAI-2000 score only	210	1.9 (1.2, 3.1)	2.1 (1.4, 3.0)
Decrease in Classic BILAG score only	164	0.9 (0.5, 1.5)	1.2 (0.8, 1.7)
Decrease in both SLEDAI-2000 and Classic BILAG scores	180	0.7 (0.4, 1.4)	2.5 (1.7, 3.7)

Table 6-37: Relationship between change in SLEDAI-2000 score (minimal change of 3) and change in Classic BILAG score in explaining change in therapy (n=1414).

Change in Score	Number of Observation	Increase in Therapy Odds Ratio (95% CI)	Decrease in Therapy Odds Ratio (95% CI)
Increase (≥ 3) in SLEDAI-2000 score only	64	8.3 (4.3, 16.2)	1.8 (0.9, 3.9)
Increase in Classic BILAG score only (exclude D to C)	119	8.5 (5.3, 13.6)	0.8 (0.4, 1.5)
Increase in both Classic BILAG and SLEDAI-2000 scores	78	25.1 (11.4, 54.8)	1.6 (0.6, 4.2)
Decrease (≥ 3) in SLEDAI-2000 score only	81	2.3 (1.2, 4.6)	1.9 (1.1, 3.2)
Decrease in Classic BILAG score only	223	0.9 (0.5, 1.4)	1.3 (1.0, 1.8)
Decrease in both SLEDAI-2000 and Classic BILAG scores	121	0.5 (0.2, 1.2)	2.1 (1.4, 3.2)

Table 6-38: Relationship between change in SLEDAI-2000 score (minimal change of 4) and change in Classic BILAG score in explaining change in therapy (n=1414).

Change in Score	Number of Observation	Increase in Therapy Odds Ratio (95% CI)	Decrease in Therapy Odds Ratio (95% CI)
Increase (≥ 4) in SLEDAI-2000 score only	57	6.8 (3.5, 13.3)	1.5 (0.7, 3.4)
Increase in Classic BILAG score only (exclude D to C)	125	8.2 (5.2, 13.1)	0.8 (0.5, 1.5)
Increase in both Classic BILAG and SLEDAI-2000 scores	72	25.2 (10.9, 58.3)	1.6 (0.6, 4.4)
Decrease (≥ 4) in SLEDAI-2000 score only	74	1.8 (0.9, 3.7)	1.8 (1.0, 3.2)
Decrease in Classic BILAG score only	233	0.9 (0.5, 1.4)	1.4 (1.0, 1.9)
Decrease in both SLEDAI-2000 and Classic BILAG scores	111	0.5 (0.2, 1.1)	2.1 (1.4, 3.2)

6.7 Discussion

This large multi-centre prospective longitudinal study had demonstrated that the BILAG-2004 index was sensitive to change as changes in the score correlated well with the corresponding change in therapy. The analysis on the subgroups of changes in overall BILAG-2004 score showed that there was a differential effect of the different subgroups of changes in its association with change in therapy. This effect was hierarchical with greater change in categories having greater predictive power which indicated that the changes in the scoring of the index were performing as expected.

The robust external responsiveness method was used in this study to assess the sensitivity to change of the disease activity indices. This method examines how the changes in the score of the index over time, relate to the corresponding changes in an external reference measure. It is imperative that the selected external reference measure represents an accepted indicator of change in the patient's status. As such, change in therapy was chosen as the external reference for disease activity for this study, in the absence of a gold standard. Therefore, the clinically meaningful change of the indices was studied. Physician's global assessment (PGA) could have been used, but this had been shown to have unsatisfactory performance with poor agreement between physicians in several studies (Gladman et al. 1994; Guzman et al. 1992; Gladman et al. 1992). It is important that the change in the index over time reflects actual change in the clinical status of the patient when the index is used to measure transition in clinical states, and this affects the number of patients that are required to detect significant differences between treatment arms in a clinical trial. Hence, the results of this study are highly relevant and applicable to clinical practice and longitudinal studies of SLE disease activity.

Although the BILAG-2004 and Classic BILAG indices were based on the principle of the physician's intention to treat, using change in therapy as the external reference should not explicitly bias the analysis in favour of these two indices as change in treatment did not determine the scoring. Only the presence of manifestations of active disease would determine the scoring. Furthermore, the scoring of these two indices was not available to the physician when the treatment decision was made and it was difficult to work out the scoring of the BILAG-2004 and Classic BILAG indices in routine clinical practice without the appropriate reference documentations.

This study also showed that Classic BILAG and SLEDAI-2000 indices were sensitive to change. In fact, this is the first study to assess the sensitivity to change of SLEDAI-2000 index since it was revised from SLEDAI index. This result is consistent with previous studies which showed that both Classic BILAG and SLEDAI indices are sensitive to change (Chang et al. 2002a; Fortin et al. 2000; Ward et al. 2000; Brunner et al. 1999; Gladman et al. 1994; Liang et al. 1989).

Although the optimal minimal change in SLEDAI-2000 score that is associated with change in therapy is 1 or 2, a higher threshold for increase in score of 3 or 4 is recommended as it is better at predicting increase in therapy (due to its superior PPV) but with a resultant decrease in the sensitivity. This is similar to the results of two previous works with SLEDAI index (Gladman et al. 2000b; Petri et al. 1991) but the study by Liang et al with SLEDAI index suggested a much higher cut-off (increase in score of at least 8 for flare of disease activity and decrease in score of at least 6 for improvement in disease activity) (Liang et al. 2004). This discrepancy could be explained by the different study design employed, with the study by Liang et al being an internet survey of lupus experts using abstracted case histories while this study is based on actual clinical practice.

Interestingly, change in SLEDAI-2000 score on its own is inadequate in explaining change in therapy. The score from which it has changed from is also equally important. This is not surprising as the score from the previous assessment put the change in score into context, for example a change from score of 18 to 15 may not constitute significant change as the patient continues to have very active disease, whereas a change of similar magnitude from 6 to 3 would most likely indicate improvement in disease activity. The result suggests that analysing change in SLEDAI-2000 score as a single cut-off would be insufficient. This has led to the use of alternative models of change in SLEDAI-2000 scores in the analysis. The

resultant alternative models showed some improvement in the performance. Nevertheless, the best model to explain change in therapy is with change in SLEDAI-2000 score (as a continuous variable) and the SLEDAI-2000 score of the previous visit included.

The comparison of the three indices did not show that any of the three indices was clearly superior with regards to sensitivity to change. The BILAG-2004 index had similar performance to Classic BILAG index. It would have been expected that the changes from Classic BILAG index to the BILAG-2004 index would have made the BILAG-2004 index more sensitive to change. However, this could not be demonstrated as the main difference between the indices was the addition of gastro-intestinal and ophthalmic manifestations of active disease. Unfortunately, these manifestations were uncommon in this study with only 13 assessments having some activity in Gastro-intestinal system and only 19 assessments with activity in Ophthalmic system. The other changes would not have made a big impact in the sensitivity to change analysis of the index and these were:

1. Improvement in terminology, definitions and glossary of the index
2. Refinement in the scoring scheme whereby improvement from Grade A reduces the score to Grade B instead of Grade C

The results of the comparison between Classic BILAG and SLEDAI-2000 indices were in contrast to previous studies which compared Classic BILAG index with SLEDAI index that showed Classic BILAG index to be superior (Ward et al. 2000; Brunner et al. 1999; Liang et al. 1989). The most likely reason for this is the different statistical methods used with external responsiveness being used in this study while previous works have used methods of internal responsiveness. External responsiveness analysis is a much more robust method as it not only determines if the index changes over time but also relates it to the corresponding change in an external reference (Husted et al. 2000). With the external reference being change

in therapy, it is the clinically meaningful change of the index over time that is being assessed. On the other hand, internal responsiveness methods (such as effect size, standardized response means and Guyatt's responsiveness index) that have been used in previous studies only assess the ability of the index to change over time. As they do not relate to change in an external reference, significant change in the index may occur without corresponding change in the clinical status. Furthermore, comparison across studies is difficult with internal responsiveness analysis as the statistics used are highly dependent on study design. This is not an issue with external responsiveness method as the statistics used are independent of study design, making the results generalisable across studies which allows for comparison (Husted et al. 2000). Apart from that, this study has involved far more patients and assessments than previous studies (less than 40 patients in all of them). Therefore, the result of the comparison of the sensitivity to change performance of the three indices in this study is highly relevant to clinical practice.

In view of the non-overlapping relationship with change in therapy when SLEDAI-2000 index was compared with the BILAG-2004 and Classic BILAG indices, further analysis was performed to define this relationship using different cut-offs in the definition of change in SLEDAI-2000 score. When used in combination with either the BILAG-2004 or Classic BILAG indices, change in SLEDAI-2000 score with single cut-off of minimal increase of 3 and minimal decrease of 1 had the best performance. This combination may be useful in the analysis of longitudinal outcome studies of SLE when both SLEDAI-2000 and the BILAG-2004 or Classic BILAG indices are used.

For the purpose of validation, overall disease activity with the BILAG-2004 and Classic BILAG indices was used, as represented by the highest score achieved by any system within the respective index (overall score). This is logical as any patient with any system

scoring a Grade A or B should be categorised as having active disease (requiring therapy in principle) regardless of how many systems have a score of A or B. However, this may put the BILAG-2004 and Classic BILAG indices at a disadvantage in the analysis due to the ceiling effect that this imposes and may underestimate the severity of the illness. For example, a patient with 3 systems scoring Grade B (and no system scoring Grade A) on one assessment which subsequently improved to just one system scoring Grade B (and no system scoring Grade A) in the next assessment would be considered as having no change in activity (and not as improved) between the two assessments, with this analysis. Further study would be required to assess the impact of this ceiling effect, using the changes in the score of individual systems in the BILAG-2004 and Classic BILAG indices in the analysis. Nevertheless, it will be important to consider the changes in the score of individual systems during data analysis in the setting of clinical trials and outcome studies.

A few other issues with the face and content validity of the index were identified during the course of this study. These were:

1. The definition of myositis and its differentiation between severe and mild based on Bohan and Peter's criteria (diagnostic criteria for polymyositis/dermatomyositis) was not appropriate. A definition based on serum muscle enzymes, muscle weakness and myalgia would be more suitable.
2. There was a potential for recording error with the item pleuropericardial pain as it could be recorded in patients with pleuritic-like chest pain, in which the physician was not convinced it was pleurisy. In order to avoid this, it was recommended that the terminology be changed to pleurisy/pericarditis and the definition to be updated to include only convincing features that the physician would consider treating and not to record if the physician was uncertain.

3. Over-scoring of Mucocutaneous system was noted with malar rash. As transient malar rash that resolved after a few days would usually require no intervention, it was recommended that for malar rash to be recorded, it needed to have been observed by a physician and to have persisted for at least one week before it could be considered sufficiently significant to be recorded.
4. An additional criterion for improvement was recommended to include manifestations that had completely resolved for at least 1 week.
5. One of the Renal system Grade A criterion for deteriorating renal function was 'Glomeular filtration rate (GFR) having fallen to < 67% of previous value' and it was noted that this could be achieved even when the GFR was within normal range. This would potentially lead to a score of at least Grade B which is inappropriate as such a change in the GFR does not reflect significant disease activity. Indeed, none of the assessment in this study with GFR that fell to less than 67% of previous value but was still within normal range was attributed to disease activity. To avoid the possibility of over-scoring of the Renal system, it was recommended that this criterion be changed to require that the GFR not only be less than 67% of previous value but that it should also be below the normal range (below 80 ml/min per 1.73 m^2).

These changes were agreed by consensus of members of BILAG and were incorporated into the index after the completion of this longitudinal study (Appendix 7). Although most of the changes were not taken into account in the present analysis, these changes would not have a major effect on the results of this study but might possibly improve on it.

In conclusion, the BILAG-2004, Classic BILAG and SLEDAI-2000 indices are sensitive to change and are suitable for use in longitudinal studies of SLE. As SLEDAI-2000 index has non-overlapping relationship with the BILAG-2004 and Classic BILAG indices

respectively, it may complement the BILAG-2004 or Classic BILAG indices in the analysis of SLE disease activity in a longitudinal study.

Chapter 7 Rasch Analysis

7.1 Abstract

Objective:

To assess the BILAG-2004 index empirically using the Rasch model

Methods:

SLE patients were recruited into a multi-centre longitudinal study and disease activity was assessed using the BILAG-2004 index. Grades D and E were combined for the analysis as both indicate inactivity, hence this index has 9 systems (items) with 4 possible scores (A, B, C or D) for each item. Analysis was with the Rasch partial credit model using RUMM2020. Fit to the model using chi-square statistic and differential item functioning (DIF) with regards to sex, race, age and disease duration using analysis of variance were assessed. DIF is a form of bias in which different groups within the sample respond differently to an individual item, given the same level of disease activity. Significance level for DIF was set at 0.0003 after Bonferroni correction. Unidimensionality was confirmed by principal component analysis (PCA) of the residuals. Assessments from the first visit were used for primary analysis while those from the second visit were used as the validation sample.

Results:

A total of 353 SLE patients were recruited (93% females, 58% Caucasian, 20% Afro-Caribbean, 19% South Asian). Rescoring of all the systems in the index (items) was required due to clustering of inactivity in all the systems, with low frequency of active disease scores (A or B). Most of the rescoring involved combining Grade A with B and Grade C with D/E. After rescoring, there was a good fit to the Rasch model: mean item fit residual -1.1 (SD 1.4),

mean person fit residual -0.3 (SD 0.4) and non-significant item-trait interaction (p=0.011). There was no significant DIF with regards to age, sex, ethnicity and disease duration. Person separation index (ability of the index to discriminate patients) was low at 0.36. Unidimensionality was confirmed by PCA with non-significant independent t-test of the residuals. Analysis of the validation sample was consistent with the primary analysis.

Conclusions:

The BILAG-2004 index, with some modification to the scoring, fits the Rasch model. This indicates that the index is a unidimensional ordinal scale index with internal construct validity. The low person separation index is consistent with clustering of patients with minimal disease activity (Grades C or D or E).

7.2 Objective

To assess the BILAG-2004 index empirically using the Rasch model.

7.3 Introduction

The Rasch model is a probabilistic mathematical model that is used to assess measurement scale empirically (Andrich 1988; Rasch 1980). It assumes that the probability of a particular response in an item is a logistic function of the severity or difficulty of the item and the patient's disease activity. Its most basic model is the dichotomous model as represented by the following equation:

$$\text{logit} = \ln\left(\frac{P}{1-P}\right) = \beta - \delta$$

where P is the probability of a patient's given response to an item, β is the patient's disease activity and δ is the severity of the item. An implicit assumption of this model is that there is a hierarchical ordering of the items in the index which means that the data forms an ordinal scale. This model has the properties of fundamental measurement which are 'additivity' of measurement units (logit), and specific objectivity as the equation allows for item severity and patient's disease activity to be independent of each other (item-person separation). Where data fit model assumptions and expectations, it also confers unidimensionality whereby the index measures only one attribute at a time. As such, this model enables assessment of internal construct validity of the index as unidimensionality infers that the items in the index are measuring the same attribute.

In this chapter, data from the longitudinal study was used to determine if the BILAG-2004 index fits the Rasch model.

7.4 Methods

The dataset from the longitudinal study (sensitivity to change study) was used in this analysis. This was a multi-centre longitudinal study in which recruited SLE patients were followed up prospectively and data were collected for all consecutive visits or encounters (inpatient or outpatient) that the patients had with the physician, as described in Chapter 6. Data were collected on disease activity as assessed using the BILAG-2004 index.

7.5 Analysis

Analysis was with Rasch partial credit model using RUMM2020 software (RUMM Laboratory, Perth, Australia). For the purpose of this analysis, BILAG-2004 Grades D and E were combined as both indicate inactivity. Therefore, the BILAG-2004 index had 9 items (representing the 9 systems in the index) with four possible categorical scores (A, B, C and D) for each item.

The fit of the data to the model was assessed by determining the overall fit of the scale and the individual item fit. The overall fit of the scale was given by the item-trait interaction which was reported as a chi-square statistic and this provided an indication of the deviation of the data from the Rasch model. An insignificant chi-square statistic would indicate that the data met model expectations and the items in the scale fit together to form a hierarchical and unidimensional scale.

Individual item fit was assessed by a number of indicators which included threshold ordering, item fit residuals and a chi-square test. Threshold refers to the point between two adjacent response categories where the probability of either response is equal. It would be expected that the thresholds would increase systematically in a logical progression from

Grade D to Grade A. Disordered thresholds occur when the scoring categories are not progressing in a logical order. The item fit residual is the standardised deviation from the model for the item and the desired range for the residual is that of between ± 2.5. Values outside this range would suggest misfit to the model. The fit of the item is further examined by a chi-square test, with a significant test indicating misfit of the item to the model expectations.

Differential item functioning (DIF) with regards to person factors, specifically sex, ethnicity (Caucasian versus non-Caucasian), age (age below 40 years versus 40 years and above) and disease duration (below 7 years versus 7 years and above), was also assessed as this could affect item fit. DIF is a form of bias in which different groups within the sample respond differently to an individual item, given the same level of the underlying attribute. For example, non-Caucasians may be more likely to have severe renal involvement compared to Caucasians, given the same level of disease activity. There are two forms of DIF, uniform and non-uniform. Uniform DIF is where there is a consistent systematic difference in the response between groups across the range of disease activity being measured. This difference varies in non-uniform DIF. Analysis of variance was used to determine the presence of DIF.

Apart from the fit statistics above, the reliability of the scores of the index (Person Separation Index) was estimated, and this provided an indication of the ability of the index to discriminate amongst the patients. This is based on Cronbach's α estimation but instead of using the raw scores, the estimates on the logit scale for each patient were used to calculate the reliability. An acceptable value for person separation index is that of at least 0.70 (Nunnally et al. 1994).

In the analysis, the patients were grouped by their level of trait, which in this study was disease activity. Given the large sample size, the number of class intervals was set to 10

for the analysis. Significance level was set at 0.01 but for DIF this was set at 0.0003 after Bonferroni correction. A test of unidimensionality utilising principal component analysis (PCA) of the residuals with independent t-test, was used.

Primary analysis was with assessments from the first visit. Confirmation of this result was made using the validation sample that comprised assessments from the second visit.

7.6 Results

A total of 353 SLE patients were recruited into this study. The demographics of the patients are summarised in Table 7-1. There were 353 assessments from the first visit that were available for the primary analysis. The distribution of disease activity according to the BILAG-2004 index across the 9 systems is summarised in Table 7-2.

Table 7-1: Demographics of patients recruited into the study (n=353).

Patient Characteristics	
Female sex (%)	92.9
Mean age in years (SD)	40.9 (12.9)
Ethnicity (%)	
Caucasian	58.1
Afro-Caribbean	20.1
South Asian	19.0
Oriental	1.4
Others	1.4
Mean disease duration in years (SD)	8.3 (7.9)

Table 7-2: Distribution of disease activity across 9 systems according to the BILAG-2004 index (n=353).

System	BILAG-2004 Score			
	A	B	C	D or E
Constitutional	2	11	18	322
Mucocutaneous	7	64	82	200
Neuropsychiatric	5	4	1	343
Musculoskeletal	10	37	109	197
Cardiorespiratory	4	13	13	323
Gastro-intestinal	0	0	3	350
Ophthalmic	0	2	1	350
Renal	11	11	10	321
Haematological	5	9	154	185

Initial Fit of Data

There were 66 patients with minimum or extreme scores (Grade D or E in all systems). Initial analysis revealed that the data did not fit the Rasch model with a mean item fit residual of -1.668 (SD 1.232), mean person location of -2.452, mean person fit residual of -0.369 (SD 0.416) and significant item-trait interaction ($p<0.001$). Individual item fit is summarised in Table 7-3. Person separation index was low at 0.46.

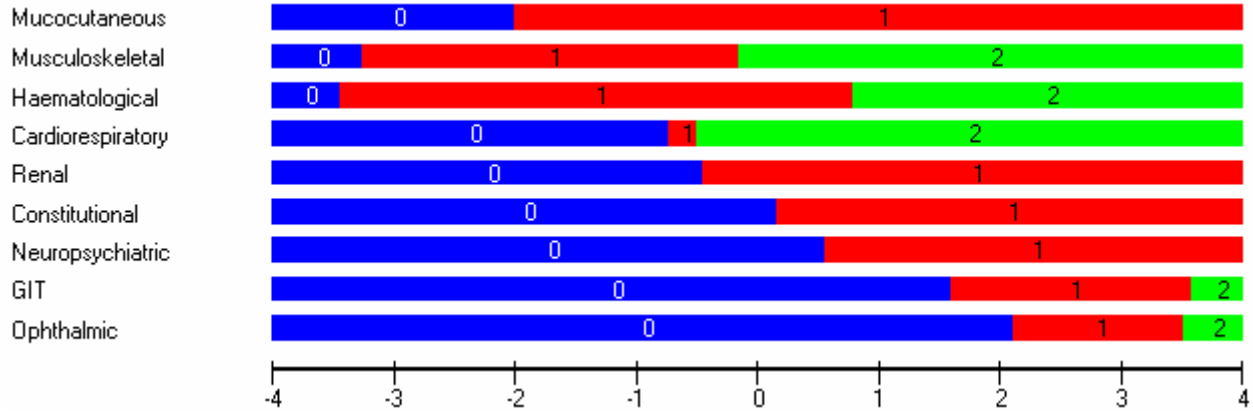
Table 7-3: Individual Item Fit of the BILAG-2004 Index with initial Rasch analysis.

System	Fit Residual	P value
Constitutional	-4.213	<0.001
Mucocutaneous	-1.371	0.001
Neuropsychiatric	-1.920	0.391
Musculoskeletal	-1.256	0.013
Cardiorespiratory	-2.056	0.276
Gastrointestinal	-0.538	0.877
Ophthalmic	-0.602	0.867
Renal	-2.710	0.016
Haematological	-0.371	0.002

Threshold Ordering

There were disordered threshold in all systems except Musculoskeletal system. The clustering of patients with Grade D or E in all the systems was contributory, necessitating grades with low frequency to be combined with the adjacent grade (Table 7-4). Threshold ordering was corrected with this rescoring (Figure 7-1). Despite having ordered threshold, Musculoskeletal system had to be rescored as it did not fit into the model with the original scoring scheme.

Figure 7-1: Threshold Map of Items in the BILAG-2004 Index after rescoring.



* items sorted according to location order with the easiest item to achieve at the top and the most difficult item to achieve at the bottom

Differential Item Functioning (DIF)

There was no DIF with regards to sex, ethnicity, age and disease duration.

Final Fit of Data

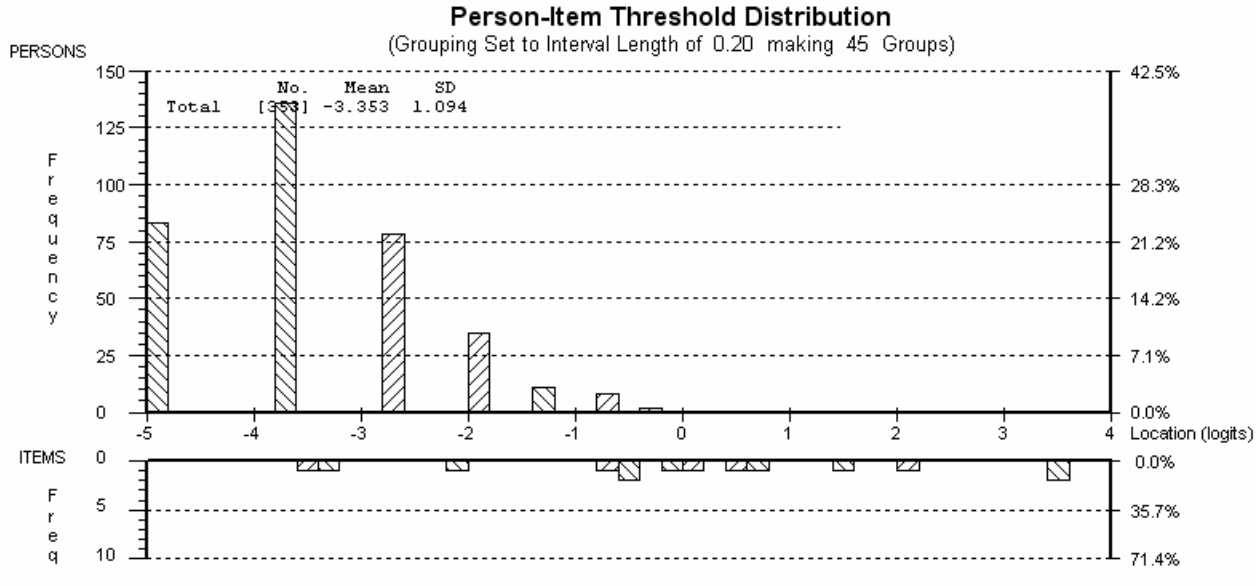
There was a good fit to the Rasch model with a mean item fit residual of -1.108 (SD 1.421), mean person location of -3.353, mean person fit residual of -0.336 (SD 0.428) and non-significant item-trait interaction (p=0.011). Person separation index remained low at 0.36. Unidimensionality was confirmed by PCA with non-significant independent t-test of the residuals. Final individual item fit is summarised in Table 7-4.

The person-item threshold distribution graph (Figure 7-2) shows the patient's disease severity and item difficulty level on the same linear (logit) scale. This indicates that the index is not appropriately targeted at the population being assessed as the majority of the patients are located towards the lower end (less disease activity) of the scale.

Table 7-4:Final Individual Item Fit of the BILAG-2004 Index.

System	Rescoring	Fit Residual	P value
Constitutional	A and B combined C and D combined	-3.274	0.017
Mucocutaneous	A and B combined C and D combined	0.603	0.078
Neuropsychiatric	A and B combined C and D combined	-1.504	0.494
Musculoskeletal	B and C combined	-0.181	0.772
Cardiorespiratory	B and C combined	-2.094	0.018
Gastrointestinal	B and C combined	-0.986	0.807
Ophthalmic	C and D combined	-0.514	0.736
Renal	A and B combined C and D combined	-2.769	0.011
Haematological	B and C combined	0.742	0.476

Figure 7-2: Person-Item threshold distribution of the BILAG-2004 index with Rasch analysis.



Validation Analysis

A total of 347 patients had assessment on their second visit which was available for this validation analysis. The distribution of disease activity according to the BILAG-2004 index across the 9 systems in this validation sample is summarised in Table 7-5.

Table 7-5: Distribution of disease activity across 9 systems according to the BILAG-2004 index in the validation sample (n=347).

System	BILAG-2004 Score			
	A	B	C	D or E
Constitutional	2	1	13	331
Mucocutaneous	5	44	75	223
Neuropsychiatric	1	1	3	342
Musculoskeletal	6	25	90	226
Cardiorespiratory	1	12	5	329
Gastro-intestinal	0	3	0	344
Ophthalmic	0	1	4	342
Renal	5	14	9	319
Haematological	2	8	164	173

Using the rescoring scheme from the primary analysis, there was no disordered threshold or DIF with regards to sex, ethnicity, age and disease duration. There was a good fit to the Rasch model with a mean item fit residual of -1.161 (SD 1.115), mean person location of -3.956, mean person fit residual of -0.326 (SD 0.357) and non-significant item-trait interaction (p=0.016). Person separation index was low at 0.32. Unidimensionality was confirmed by PCA with non-significant independent t-test of the residuals. The individual item fit is summarised in Table 7-6. This result is consistent with the results from the primary analysis and hence, validates the primary results.

Table 7-6: Individual Item Fit of the BILAG-2004 Index with the validation sample.

System	Fit Residual	P value
Constitutional	-3.274	0.017
Mucocutaneous	0.603	0.078
Neuropsychiatric	-1.504	0.494
Musculoskeletal	-0.181	0.772
Cardiorespiratory	-2.094	0.018
Gastrointestinal	-0.986	0.807
Ophthalmic	-0.514	0.736
Renal	-2.769	0.011
Haematological	0.742	0.476

7.7 Discussion

The results of this study showed that the BILAG-2004 index, with some modifications to the scoring, had a good fit to the Rasch model. This confirms that the index is performing as an ordinal scale index with hierarchical ordering of the system scores, in the way it is designed. It also indicates that this index is unidimensional, in which all the system scores are

measuring a single latent trait (that is disease activity), and has internal construct validity whereby all the system scores correlate well with each other in measuring disease activity.

It is apparent that there is a mismatch between the difficulty of the index and disease activity of the population under study as indicated by the mean patient location which is well below zero, and the person-item threshold distribution graph (Figure 7-2). This indicates that the index is designed to capture much more severe disease activity (hence much more difficult items in the index to achieve) than the population studied. This is not unexpected as many of the patients have low level of disease activity. Even in patients with active disease, most have only one system involved while the other systems are usually unaffected or minimally affected.

It is the clustering of patients with minimal disease activity that has resulted in low patient separation index (reliability) whereby the BILAG-2004 index is not able to differentiate between patients within this group. However, differentiation within this group of patients with minimal disease activity is not clinically important as it would have little or no impact on the management of the disease. Therefore, although the low patient separation index is not ideal, it is acceptable considering the circumstances.

In the analysis, all of the systems in the index required rescoring to enable the index to fit to the Rasch model. Once again, this is most likely related to the clustering of Grade D or E scores in all the systems while there was low frequency of Grade A and B scores. As a result, the grades with low frequencies had to be combined with the adjacent grade. It is reassuring that the rescoring scheme is relevant from clinical viewpoint as the majority of combinations of categories are that of active disease (combination of Grades A and B) or minimal disease activity (combination of Grades C and D). As the scoring in all the systems are not uniform following rescoring, linear transformation of the ordinal scale using the Rasch model is not

suitable for use outside of this study. If linear transformation using the Rasch model that is generally applicable is required, a uniform scoring across all the system is essential. This would involve a study with targeted recruitment of patients with active disease, hence increasing the frequency of Grades A and B, and reducing the clustering of patients in Grades D or E.

Chapter 8 Final Discussion

The development and validation process for the BILAG-2004 index had led to major changes in the content of the index and its supporting documentation as compared to its predecessor, the Classic BILAG index. The major improvement had been in the terminology and definitions of most items, which was reflected in the glossary of 18 pages for the BILAG-2004 index (Appendix 7). In contrast, the Classic BILAG index had a 2-page glossary and as such, relied mainly on the judgement of the individual physician who was completing the index (Appendix 2). This may have resulted in less than optimal standardisation across different physicians and introduce user-related errors in studies involving more than one physician. This issue was recognised and for clinical trials, a longer and more detailed glossary had been developed for the Classic BILAG index, whereas the detailed glossary for the BILAG-2004 index was developed from the start, with the aim of ensuring greater standardisation amongst users of the index. Explicit and clear definitions/qualifications for the majority of the items in the index were incorporated. In addition, overly subjective definitions were avoided and at the same time, definitions were checked to ensure that they were not too restrictive or stringent so that the index remained comprehensive (able to capture all aspects of disease activity and severity). Items that were difficult to attribute to disease activity with certainty such as fatigue, mood disorder, anxiety disorder and migraine/tension/cluster headache, were removed. Despite these changes, the BILAG-2004 index still requires an element of physician's subjective judgement in deciding the attribution to SLE disease activity for certain manifestations that are difficult to differentiate from other conditions, such as lupus pancreatitis, lupus cholecystitis, cognitive dysfunction and lupus abdominal peritonitis/serositis.

This research on the validation of the BILAG-2004 index represents the largest multi-centre validation study to-date of a clinical disease activity index in SLE. The major strength of these validation studies is in the involvement of several physicians (from different centres) who have recruited a large number of patients (and assessments). Furthermore, the data were collected prospectively in routine clinical setting that should be applicable to routine practice and clinical trials. Previous validation studies of other disease activity indices have involved rather small sample sizes (usually less than 50 patients) and/or retrospective assessment or abstracted case histories (Gladman et al. 2002; Bae et al. 2001; Fortin et al. 2000; Ward et al. 2000; Brunner et al. 1999; Gladman et al. 1994; Hay et al. 1993; Bombardier et al. 1992; Petri et al. 1992; Gladman et al. 1992; Liang et al. 1989).

This index had undergone a thorough examination, with robust statistical methods being used in the analysis. The reference standard for disease activity of change in therapy used in the criterion validity and sensitivity to change analyses might not be perfect but remains the best benchmark. This was chosen in preference to physician's global assessment of disease activity as this had been shown in several studies to have unsatisfactory performance and poor agreement between expert physicians, making it unsuitable for such a purpose (Wollaston et al. 2004; Brunner et al. 1999; Gladman et al. 1994; Gladman et al. 1992). The definition used for change in therapy in this research was meticulous and robust. There was no issue with circularity even though the index was based on the principle of the physicians' intention to treat, as the treatment was not taken into account in the scoring. Furthermore, the treatment decisions made in the clinic were not based on the score of the index, as the score was not available to the physician when the treatment decisions were made. Most of the participating physicians were not familiar with the scoring scheme, although some of them had helped in the initial development of the index (during the

consensus building process). With change in therapy as the external standard, the results of the criterion validity and sensitivity to change analyses are clinically meaningful. Therefore, the results from this validation process will be highly relevant and applicable to clinical practice.

There is a possibility that the effect of using these indices may have influenced and changed the way in which the physicians assessed the patients. Physicians have to be thorough in their assessment and this may have been much more meticulous than in routine practice. As a result, subtle changes in the patients' status would be detected earlier leading to change in treatment (occurring earlier than would have been expected in routine practice). Although it is difficult to prove, this positive effect is one that is welcomed and should not have an effect on the results as the trigger to a change in treatment is the detection of change in clinical status rather than completion of the indices.

The reliability study has clearly demonstrated the importance of training before physicians start using a composite clinical disease activity index. It must be emphasised that one cannot assume that physicians, without any form of training, will use composite clinical disease activity indices as designed. Training can be considered to be equivalent to standardisation of equipment whereby the equipment in this setting is the combination of the user (physician) and the disease activity index. It is important to realise that how the index is completed will affect the result and if done incorrectly, will lead to unacceptable user-related error which cannot be controlled statistically. This is a significant issue in this complex multi-system disease that requires multiple domains to be considered (such as disease activity, damage, health status and intercurrent illness), but only features attributable to SLE disease activity are to be recorded in the index. The reliability study highlighted this issue of attribution in the first exercise which was significantly reduced with training prior to the

second exercise. Therefore, the aims of the training are twofold. Firstly, it is to familiarise the physician (user) with the index and the definitions used in the index (not only of the items in the index but also what is meant by 'new', 'improved', 'worse' and 'unchanged') so that the physician will be competent in completing the index correctly. In addition, the training is to increase the awareness of the physician on the issue of attribution of manifestations when assessing SLE patients and to record only manifestations that are due to SLE disease activity in the index. Apart from that, it is essential for trial monitors to be vigilant and meticulous on this issue of attribution in clinical trials setting. In the reliability study, the training was not performed in a formal setting due to logistic reasons. However, formal training with the use of paper cases would be recommended in future studies involving the BILAG-2004 index. The pharmaceutical companies that are engaged in current clinical trials in SLE are aware of the importance of this issue and have organised formal training for investigators prior to the start of the trial.

One of the limitations in the analyses is with the use of overall score for the BILAG-2004 index to represent overall disease activity, in the cross sectional and sensitivity to change studies. The overall score for the index was determined by the highest score achieved by any system within the index. The use of this overall score is logical and sufficient for the purpose of validation. However, it possibly have under-estimated the ability of the index due to the ceiling effect imposed by the overall score and this may have an impact in the analyses comparing the performance of the three indices (the BILAG-2004, Classic BILAG and SLEDAI-2000 indices). Therefore, further analysis using all the individual system scores is warranted.

In the cross sectional and longitudinal (sensitivity to change) studies, all patients were assessed simultaneously using the BILAG-2004, Classic BILAG and SLEDAI-2000 indices.

Although these three indices have similar items for several manifestations of active disease, there are differences in the definitions used across the indices. Hence, there is a possibility that the same definition had been used unintentionally across the 3 indices for similar items, but this is difficult to prove. This is especially so with the BILAG-2004 and Classic BILAG indices which are quite similar. An example of this is with inflammatory arthritis, in which the definition according to the three indices is summarised in Table 8-1.

Table 8-1: Definition for inflammatory arthritis according to the BILAG-2004, Classic BILAG and SLEDAI-2000 indices.

Index	Definition for inflammatory arthritis
SLEDAI-2000	≥ 2 joints with pain and signs of inflammation (tenderness, swelling or effusion)
BILAG-2004	(a) Severe arthritis Observed active synovitis ≥ 2 joints with marked loss of functional range of movements and significant impairment of activities of daily living, that has been present on several days (cumulatively) over the last 4 weeks (b) Moderate arthritis or Tendonitis or Tenosynovitis Tendonitis/tenosynovitis or active synovitis ≥ 1 joint (observed or through history) with some loss of functional range of movements, that has been present on several days over the last 4 weeks (c) Mild arthritis or Arthralgia or Myalgia Inflammatory type of pain (worse in the morning with stiffness, usually improves with activity & not brought on by activity) over joints/muscle and/or inflammatory arthritis which does not fulfil the above criteria for moderate or severe arthritis
Classic BILAG	(a) Severe polyarthritis with loss of function Active joint inflammation with clinically significant loss of the functional range of movement of the involved joints. (b) Arthritis Active joint inflammation without loss of functional range of motion (c) Arthralgia

Both the BILAG-2004 and Classic BILAG indices have three items for inflammatory arthritis, dividing it into three levels of severity (mild, moderate and severe) while SLEDAI-2000 has

only one item for the same manifestation. It is apparent that the criteria used by the three indices for inflammatory is quite different. Even though both the Classic BILAG and BILAG-2004 indices have similar division of inflammatory arthritis into three levels of severity, what could be recorded as severe arthritis in Classic BILAG index may only be recorded as moderate arthritis in the BILAG-2004 index, due to the different criteria used in the definition. Similarly, the same definitions for improvement in a manifestation and for new manifestation may have been used for both the BILAG-2004 and Classic BILAG indices. The improvement criteria in the BILAG-2004 index incorporate 2 parameters which include the amount of improvement and the duration of improvement, whereas there is no specific definition for improvement in Classic BILAG index (it is based solely on the physician's judgement). There is also a slight difference in the definition of new manifestation between the BILAG-2004 and Classic BILAG indices, in which new manifestation that has also improved is recorded as improved in the BILAG-2004 index (but not in Classic BILAG index) and this difference could be overlooked. One possible way around this problem of using the same definition across the indices is for different indices to be completed by different physicians. However, this will require a patient to be seen by 3 physicians at every assessment, which is not practical or feasible in a large multi-centre study such as the cross sectional and sensitivity to change studies.

This index has not been tested through its full range due to clustering of observations with low level of disease activity, which has been highlighted in the Rasch analysis. This is particularly evident in the Gastrointestinal and Ophthalmic systems. A better distribution of disease activity with more Grades A and B in all of the systems would have been ideal, but this has proved difficult to achieve despite recruitment being targeted at patients with active disease. Indeed, manifestations of active disease in the Gastrointestinal and Ophthalmic

systems were rare in the cross sectional and longitudinal studies. However, these manifestations are associated with significant morbidity and mortality (requiring change in therapy), hence it is important to be able to record them whenever they occur.

Although the validation studies had a good case-mix of patients with broad representation of the three main ethnic groups in United Kingdom (Caucasians, Afro-Caribbeans and South Asians), this index had not been used in other ethnic groups (such as Orientals and Hispanics) and in the setting outside of United Kingdom (particularly in countries where English is not the first language). It remains to be seen if the results are applicable in a different setting and to these other ethnic groups, although it is not anticipated to perform differently providing that the physicians are experienced in the management of lupus, able to speak English fluently and appropriately trained to use the index. There may be a need for translation and validation of this index in other languages to optimise its use in non-English speaking countries.

Unfortunately, the predictive validity of this index has yet to be demonstrated. A multi-centre longitudinal study designed to determine if disease activity according to the BILAG-2004 index predicts development of organ damage in an inception cohort of newly diagnosed patients, is on-going. Patients who are within 12 months of achieving the 4th ACR classification criteria of SLE are recruited and followed up prospectively. Development of damage is assessed using ACR/SLICC damage index. Recruitment into this study began in March 2005 and as of 30th April 2007, 83 patients had been recruited. However, the data currently are inadequate for any conclusive analysis to be made as far fewer patients have developed damage than expected. This reflects the improvement in the management of SLE patients, as the initial calculation of the number of patients required for this study (n=60) is based on a study that was performed in the early 1990s (Stoll et al. 1996). Clearly, more

patients and much longer follow-up are required. This study is continuing and it has been extended for another 5 years with ethical approval.

The results presented in this thesis have shown that the BILAG-2004 index is a comprehensive index that is valid for use to assess SLE disease activity. It should be suitable for use in clinical trials. The ability of this index to provide disease activity score across different systems will be useful in the assessment of therapeutic agents, in particular those that may only be helpful for disease activity in certain systems (such as musculoskeletal) but not others. This index has been well accepted by the international rheumatology community and several clinical trials in SLE are using this index as an outcome measure. Apart from clinical trials, this index will be useful in daily clinical setting. With the arrival of newer and expensive therapies, health resource planning especially budgeting for these therapies will be essential. Similar to the usage of biological therapy for rheumatoid arthritis, I envisage that standardised objective assessment of disease activity will become a requirement to demonstrate eligibility and response to these expensive therapies. It has been argued that this index may be too complicated and difficult to complete but this has not been the case from the validation studies. It only takes a few minutes to complete the index once the physician is familiar with the items in the index and the glossary definitions for them, as most patients have very few items to record and the time required to complete the index mirrors the complexity of the case. The difficult part is the scoring of the index, which in fact is the most complicated part of the index. Fortunately, a purpose-built database incorporating the BILAG-2004 index is under development, which will facilitate data management and analyses.

8.1 Future Research

As discussed earlier, the use of the overall score for both the BILAG-2004 and Classic BILAG indices in the validation analyses may have under-estimated the ability of these two indices. Apart from that, the results of the sensitivity to change analysis revealed that there was a non-overlapping relationship with change in therapy when SLEDAI-2000 index was compared with the BILAG-2004 and Classic BILAG indices. This suggests that there is a role of combining the SLEDAI-2000 with either the BILAG-2004 or Classic BILAG indices in the analysis of longitudinal outcome studies in SLE. As the analysis was undertaken using overall score for both the BILAG-2004 and Classic BILAG indices, further analysis which takes into account all of the individual system scores is required as this would provide a better understanding of how the individual system scores would relate to treatment decision and would better define the relationship between the BILAG-2004, Classic BILAG and SLEDAI-2000 indices.

Furthermore, the definition of flare based on the BILAG-2004 index has not been addressed. Flare is used in longitudinal studies to indicate deterioration or worsening of disease activity between two time points, such as a change from Grade D to A. The difficulty is in the differentiation of flare from persistent activity. For the majority of the items in the index, an improvement will result in a drop of one category in the score. If the manifestation remained unchanged subsequently, the score will increase back to its initial score, giving an impression that the patient had flared (which is incorrect as the manifestation has not worsened) when in fact this increase in score is due to persistent activity. It should be pointed out that this change in score is appropriate in reflecting the need for treatment as the manifestation is no longer improving, but it is the interpretation of the change in score that one needs to be careful with in terms of defining flare. As an illustration, a new subacute

cutaneous lupus rash affecting 10% of body surface area would result in a Mucocutaneous system score of B (Visit 1). With treatment, the rash improved to affect only 4% of body surface area and the Mucocutaneous system score dropped to Grade C (Visit 2). In the following month, if the rash remained unchanged, the Mucocutaneous system score would go back up to Grade B (Visit 3). Therefore, the Mucocutaneous system score had changed from Grade B to C between the first two visits (Visit 1 and 2) and from Grade C to B between latter two visits (Visit 2 and 3). At face value, the change in the Mucocutaneous system scores between Visit 2 and 3 will be classified incorrectly as a flare. This issue with flare definition is not unique to the BILAG-2004 index but is universal across all other indices. The beauty of being a transitional index (both the BILAG-2004 and Classic BILAG indices) is that it is able to differentiate whether a manifestation has improved, worsened or stayed unchanged, which will allow the issue of flare definition to be addressed. As the status of the manifestation ('new', 'improved', 'worsened' or 'unchanged') is recorded in the index, it can be used to identify persistent activity (whereby the manifestation has stayed unchanged) as opposed to flare (whereby the manifestation is new or has worsened). However, this is very difficult to address with non-transitional indices especially global score indices, which by design are unable to differentiate between the changing status of a manifestation (improved, worsened or unchanged). Using the earlier clinical scenario of subacute cutaneous rash, SLEDAI-2000 index score is the same (score of 2) for all three visits and similarly, the score would remain the same even if the rash has worsened to involve 40% of body surface area.

As the BILAG-2004 index is designed as an ordinal scale, it is recommended that it should be analysed as such. However, there are some restrictions to the way analysis can be performed, as there are limited statistical methods that can accommodate ordinal data. Hence, it would be useful if the BILAG-2004 index score could be converted into a continuous

numerical score for use in certain statistical circumstances where analysis as ordinal data is not feasible. This is possible using the Rasch model to transform the data into a linear scale but it would require another study, with targeted recruitment of patients with active disease that has good distribution across the different scoring categories in all the 9 systems. Another possibility is to develop an additive numerical scoring scheme by modelling change in therapy using the scores of the individual systems with regression analysis. The coefficients from the fitted model can be used to define the numerical values of each categorical score.

Finally, there is a need for a disease activity index that has been validated for use in pregnancy. In the past, pregnancy in SLE was considered to be a serious problem, not infrequently resulting in significant risk of mortality to mother and baby. Over the last 2 to 3 decades, the survival and treatment of SLE have improved considerably resulting in more SLE patients getting pregnant. This poses a special problem as pathophysiological changes of pregnancy may mimic or confound disease activity. Although several of the indices that are developed for use in non-pregnant state have been modified for use in pregnancy, none of them have been formally validated. Only the modification of LAI for use in pregnancy (LAI-P) had undergone validation in the form of criterion validity and sensitivity to change (Ruiz-Irastorza et al. 2004). This study involved 38 patients with 158 assessments and LAI-P was only compared with physician's global assessment in both the criterion validity and sensitivity to change analyses, which was not ideal due to the drawback of using physician's global assessment (discussed earlier). Apart from that, the sensitivity to change analysis used the less robust method of internal responsiveness in the form of standardised response mean (see Section 1.8.4 for more details on methods of sensitivity to change analysis). A disease activity index for use in pregnancy based on the BILAG-2004 index has been developed (BILAG2004-Pregnancy index). This has involved modifications that take into account

pathophysiological changes of pregnancy that may be confused with disease activity. This index is currently undergoing validation which commenced in February 2006 and is anticipated to be completed in late 2008.

8.2 Conclusion

In conclusion, the results of the research studies presented in this thesis have achieved most of the objectives of the thesis as set out, which is to develop and validate the BILAG-2004 index for the assessment of SLE disease activity. Only the predictive validity of this index remains outstanding, but the study is on-going.

Appendix 1 BILAG-2004 Index

BILAG-2004 INDEX ASSESSMENT

Only features attributable to SLE are to be recorded and refer to the last 4 weeks compared with the previous 4 weeks.

♦♦ THIS SHOULD BE USED WITH THE GLOSSARY ♦♦

Scoring for features: **ND Not Done**
1 Improving
2 Same
3 Worse
4 New
Yes/No OR Value
❑ indicate if feature not due to lupus
(default is 0 = not present)

CONSTITUTIONAL

1. Pyrexia (documented)	(	)
2. Weight loss - unintentional > 5%	(	)
3. Lymphadenopathy/splenomegaly	(	)
4. Fatigue/malaise/lethargy	(	)
5. Anorexia	(	)

MUCOCUTANEOUS

6. Skin eruption - severe	(	)
7. Skin eruption - mild	(	)
8. Angio-oedema - severe	(	)
9. Angio-oedema - mild	(	)
10. Mucosal ulceration - severe	(	)
11. Mucosal ulceration - mild	(	)
12. Panniculitis/Bullous lupus - severe	(	)
13. Panniculitis/Bullous lupus - mild	(	)
14. Cutaneous vasculitis/thrombosis	(	)
15. Digital infarcts/nodular vasculitis	(	)
16. Alopecia - severe	(	)
17. Alopecia - mild	(	)
18. Peri-ungual erythema/chilblains	(	)
19. Splinter haemorrhages	(	)

NEUROPSYCHIATRY

20. Aseptic meningitis	(	)
21. Cerebral vasculitis	(	)
22. Demyelinating syndrome	(	)
23. Myelopathy	(	)
24. Acute confusional state	(	)
25. Psychosis	(	)
26. Acute inflammatory demyelinating polyradiculoneuropathy	(	)
27. Mononeuropathy (single/multiplex)	(	)
28. Cranial neuropathy	(	)

29. Plexopathy ( )
30. Polyneuropathy ( )
31. Seizure disorder ( )
32. Status epilepticus ( )
33. Cerebrovascular disease (not due to vasculitis) ( )
34. Cognitive dysfunction ( )
35. Movement disorder ( )
36. Autonomic disorder ( )
37. Cerebellar ataxia ( )
38. Headache, severe, unremitting ( )
39. Headache - migraine/cluster/tension ( )
40. Headache from IC hypertension ( )
41. Mood disorder (depression/mania) ( )
42. Anxiety disorder ( )

MUSCULOSKELETAL

43. Definite myositis (Bohan & Peter) ( )
44. Myositis with incomplete criteria ( )
45. Severe polyarthritis ( )
46. Arthritis/Tendonitis ( )
47. Arthralgia/Myalgia ( )

CARDIORESPIRATORY

48. Myocarditis - mild ( )
49. Myocarditis/Endocarditis + Cardiac failure ( )
50. Arrhythmia ( )
51. New valvular dysfunction ( )
52. Serositis (pleuro-pericardial pain) - mild ( )
53. Cardiac tamponade ( )
54. Pleural effusion with dyspnoea ( )
55. Pulmonary haemorrhage/vasculitis ( )
56. Interstitial alveolitis/pneumonitis ( )
57. Shrinking lung syndrome ( )
58. Aortitis ( )
59. Coronary vasculitis ( )

GASTROINTESTINAL

60. Peritonitis ( )
61. Abdominal serositis or ascites ( )
62. Lupus enteritis/colitis ( )
63. Malabsorption ( )
64. Protein losing enteropathy ( )
65. Intestinal pseudo-obstruction ( )
66. Hepatitis ( )
67. Acute cholecystitis ( )
68. Acute pancreatitis ( )

OPHTHALMIC

69. Orbital inflammation/myositis/proptosis		( )	
70. Keratitis - severe		( )	
71. Keratitis - mild		( )	
72. Anterior uveitis		( )	
73. Posterior uveitis/retinal vasculitis - severe		( )	
74. Posterior uveitis/retinal vasculitis - mild		( )	
75. Episcleritis		( )	
76. Scleritis - severe		( )	
77. Scleritis - mild		( )	
78. Retinal/choroidal vaso-occlusive disease		( )	
79. Isolated cotton-wool spots (cytoid bodies)		( )	
80. Optic neuritis		( )	
81. Anterior ischaemic optic neuropathy		( )	

RENAL

82. Systolic blood pressure (mm Hg)	value	( )	❑
83. Diastolic blood pressure (mm Hg)	value	( )	❑
84. Accelerated hypertension	Yes/No	( )	
85. Urine dipstick	(-=0,+=1, ++=2, +++=3)	( )	❑
86. Urine albumin-creatinine ratio	mg/mmol	( )	❑
87. Urine protein-creatinine ratio	mg/mmol	( )	❑
88. 24 hour urine protein (g)	value	( )	❑
89. Nephrotic syndrome	Yes/No	( )	
90. Creatinine (plasma/serum)	µmol/l	( )	❑
91. GFR (calculated)	ml/min	( )	❑
92. Active urinary sediment	Yes/No	( )	
93. Active nephritis	Yes/No	( )	

HAEMATOLOGICAL

94. Haemoglobin (g/dl)	value	( )	❑
95. Haematocrit	value	( )	❑
96. Total white cell count (x 10^9/l)	value	( )	❑
97. Neutrophils (x 10^9/l)	value	( )	❑
98. Lymphocytes (x 10^9/l)	value	( )	❑
99. Platelets (x 10^9/l)	value	( )	❑
100. TTP		( )	
101. Evidence of active haemolysis	Yes/No	( )	
102. Coombs' test positive (isolated)		( )	

BILAG2004 INDEX GLOSSARY

- all features must be attributable to SLE and refer to the last 4 weeks compared with the previous 4 weeks

- in some manifestations, it may be difficult to differentiate SLE from other causes as there may not be any specific test and the decision would then lies with the physician's judgement on the balance of probabilities

- definition of improvement: (a) the amount of improvement is sufficient for consideration of reduction in therapy

  (b) improvement must be present ≥ 2 weeks of the previous 4 weeks

- most are self-explanatory but definitions are available for most descriptors

- ophthalmic manifestations need to be assessed by ophthalmologist

CONSTITUTIONAL

1. Pyrexia	temperature > 37.5˚C documented
2. Unintentional weight loss > 5%	
3. Lymphadenopathy	palpable lymph node more than 1 cm diameter
4. Fatigue or malaise or lethargy	
5. Anorexia	

MUCOCUTANEOUS

6. Severe eruption	> 18% body surface area includes discoid lesion body surface area (BSA) is defined using the rules of nines (used to assess extent of burns) as follows: palm (excluding fingers) = 1% BSA each lower limb = 18% BSA each upper limb = 9% BSA torso (front) = 18% BSA torso (back) = 18% BSA head = 9% BSA genital (male) = 1% BSA

7. Mild eruption	≤ 18% body surface area includes discoid lesion
8. Angio-oedema	potentially life-threatening eg: stridor
9. Severe mucosal ulceration	disabling extensive &/or deep ulceration
10. Mild mucosal ulceration	localised non-disabling ulceration
11. Severe panniculitis	any one: affecting the face > 9% body surface area threatens integrity of epithelium &/or subcutaneous tissue
12. Mild panniculitis	≤ 9% body surface area and does not fulfill any criteria for severe panniculitis
13. Cutaneous vasculitis/thrombosis	resulting in gangrene or ulceration or skin infarction
14. Digital infarct/nodular vasculitis	localised single or multiple infarct(s) over digit(s) or tender erythematous nodule(s)
15. Severe alopecia	clinically detectable diffuse or patchy hair loss with scalp inflammation
16. Mild alopecia	not clinically detectable and little/no scalp inflammation (may be diffuse & must be spontaneous)
17. Peri-ungual erythema or chilblains	
18. Splinter haemorrhages	

NEUROPSYCHIATRIC

19. Aseptic meningitis	criteria (all): acute/subacute onset headache photophobia neck stiffness fever signs of meningeal irritation abnormal CSF but negative cultures exclude CNS/meningeal infection, intracranial haemorrhage

20. Cerebral vasculitis	should be present with features of vasculitis in another system and supportive imaging &/or biopsy findings
21. Demyelinating syndrome	discrete white matter lesion with associated neurological deficit not recorded elsewhere there must have been at least one previously recorded event exclude multiple sclerosis
22. Myelopathy	acute onset of rapidly evolving paraparesis or quadriparesis and/or sensory level exclude intramedullary and extramedullary space occupying lesion
23. Acute confusional state	acute disturbance of consciousness or level of arousal with reduced ability to focus, maintain or shift attention includes hypo- and hyperaroused states and encompasses the spectrum from delirium to coma
24. Psychosis	delusion or hallucinations does not occur exclusively during course of a delirium exclude drugs, substance abuse, primary psychotic disorder
25. Acute inflammatory demyelinating polyradiculoneuropathy	criteria: progressive polyradiculoneuropathy loss of reflexes symmetrical involvement increased CSF protein without pleocytosis supportive abnormal nerve conduction study
26. Mononeuropathy (single/multiplex)	nerve conduction study not essential
27. Cranial neuropathy	except optic neuropathy which is classified elsewhere
28. Plexopathy	disorder of brachial or lumbosacral plexus

	resulting in neurological deficit not corresponding to territory of single root or nerve positive electrophysiology study required
29. Polyneuropathy	symmetrical distal sensory and/or motor deficit positive electrophysiology study required
30. Seizure disorder	independent description of seizure by reliable witness
31. Status epilepticus	a seizure or series of seizures lasting ≥ 30 minutes without full recovery to baseline
32. Cerebrovascular disease (not due to vasculitis)	any one with supporting imaging: stroke syndrome transient ischaemic attack intracranial haemorrhage exclude hypoglycaemia, cerebral sinus thrombosis, vascular malformation, tumour, abscess cerebral sinus thrombosis not included as definite thrombosis not considered part of lupus activity
33. Cognitive dysfunction	significant deficits in any cognitive functions: simple attention complex attention memory visual-spatial processing language reasoning/problem solving psychomotor speed executive functions neuropsychological testing should be done if possible or corroborating history from third party that it is interfering with daily activities exclude substance abuse
34. Movement disorder	exclude drugs
35. Autonomic disorder	any one: fall in blood pressure to standing > 30/15 mm

	Hg (systolic/diastolic)
	increase in heart rate to standing ≥ 30 bpm
	loss of heart rate variation with respiration (max – min < 15 bpm, expiration:inspiration ratio < 1.2, Valsalva ratio < 1.4)
	loss of sweating over body and limbs (anhidrosis) by sweat test
	exclude drugs and diabetes mellitus
36. Cerebellar ataxia	
37. Severe headache (unremitting)	disabling headache unresponsive to narcotic analgesia & lasting ≥ 3 days
	exclude intracranial space occupying lesion and CNS infection
38. Migraine with/without aura	recurrent attacks of headache lasting 4 - 72 hours
	may be preceded by neurological aura (lasting up to 1 hour)
39. Tension headache	recurrent episodes of headaches lasting minutes to days
40. Cluster headache	attacks of severe unilateral headache lasting 15 - 180 minutes
	attacks at least once every other day and up to 8 times a day
	attacks occur in clusters (series of weeks or months) separated by remissions of usually months or years
41. Headache from IC hypertension	exclude cerebral sinus thrombosis
42. Mood disorder (depression/mania)	prominent & persistent disturbance in mood characterised by depressed mood or markedly diminished interest or pleasure in almost all activities or elevated, expansive or irritable mood
	should result in significant distress or impaired functioning

43. Anxiety disorder	prominent anxiety, panic disorder, panic attacks or obsessions or compulsions resulting in clinically significant distress or impaired functioning
MUSCULOSKELETAL	
44. Definite myositis	≥ 3 Bohan & Peter criteria: proximal muscle weakness elevated muscle enzymes positive muscle biopsy abnormal EMG
45. Incomplete myositis	2 Bohan & Peter criteria
46. Severe polyarthritis	observed active synovitis ≥ 2 joints with significant impairment of activities of daily living and has been present on several days (cumulatively) over the last 4 weeks
47. Arthritis or Tendonitis	tendonitis or active synovitis ≥ 1 joint with some impairment of function, which has been present on several days over the last 4 weeks
48. Arthralgia or Myalgia	inflammatory joint or muscle pain which does not fulfill the above criteria for arthritis or myositis
CARDIORESPIRATORY	
49. Mild myocarditis	inflammation of myocardium with raised cardiac enzymes &/or ECG changes and without resulting cardiac failure, arrhythmia or valvular dysfunction
50. Cardiac failure	cardiac failure due to myocarditis or non-infective inflammation of endocardium or cardiac valves (endocarditis)
51. Arrhythmia	arrhythmia (except sinus tachycardia) due to myocarditis or non-infective inflammation of endocardium or cardiac valves (endocarditis)

52. New valvular dysfunction	new cardiac valvular dysfunction due to myocarditis or non-infective inflammation of endocardium or cardiac valves (endocarditis)
53. Mild serositis (pleuro-pericardial pain)	in absence of cardiac tamponade or pleural effusion with dyspnoea
54. Cardiac tamponade	
55. Pleural effusion with dyspnoea	
56. Pulmonary haemorrhage/vasculitis	inflammation of pulmonary vasculature with haemoptysis &/or dyspnoea &/or pulmonary hypertension
	supporting imaging &/or histological diagnosis
57. Interstitial alveolitis/pneumonitis	radiological features of alveolar infiltration not due to infection or haemorrhage
	reduced corrected gas transfer Kco (< 70% normal)
58. Shrinking lung syndrome	reduced lung volumes (< 70% predicted) in presence of normal corrected gas transfer Kco with dysfunctional diaphragmatic movements
59. Aortitis	inflammation of aorta with or without dissection with supporting imaging abnormalities
	accompanied by > 10 mm Hg difference in BP between arms &/or claudication of extremities &/or vascular bruits
60. Coronary vasculitis	inflammation of coronary vessels with radiographic evidence of non-atheromatous narrowing, obstruction or aneurismal changes

GASTROINTESTINAL

61. Peritonitis	serositis presenting as acute abdomen with rebound/guarding
62. Serositis	not presenting as acute abdomen
63. Lupus enteritis or colitis	vasculitis or inflammation of small or large bowel with supportive imaging &/or biopsy findings

64. Malabsorption	diarrhoea with abnormal D- xylose absorption test or increased faecal fat excretion after exclusion of coeliac's disease (poor response to gluten-free diet) and gut vasculitis
65. Protein-losing enteropathy	diarrhoea with hypoalbuminaemia or increased fecal excretion of iv radiolabeled albumin after exclusion of gut vasculitis
66. Intestinal pseudo-obstruction	subacute intestinal obstruction due to intestinal hypomotility
67. Hepatitis	raised transaminases in absence of autoantibodies specific to autoimmune hepatitis (eg: anti-smooth muscle, anti-liver cytosol 1) &/or biopsy appearance of chronic active hepatitis
68. Acute cholecystitis	after exclusion of gallstones and infection
69. Acute pancreatitis	usually associated multisystem involvement

RENAL

70. Systolic blood pressure	
71. Diastolic blood pressure	
72. Accelerated hypertension	blood pressure rising to > 170/110 mm Hg within 1 month with grade 3 or 4 Keith-Wagener-Barker retinal changes (flame-shaped haemorrhages or cotton-wool spots or papilloedema)
73. Urine dipstick	
74. Urine albumin-creatinine ratio	on freshly voided urine sample
75. Urine protein-creatinine ratio	on freshly voided urine sample
76. 24 hour urine protein	
77. Nephrotic syndrome	criteria: heavy proteinuria (> 50 mg/kg/day or > 3.5 g/day or protein-creatinine ratio > 350 mg/mmol or albumin-creatinine ratio > 350 mg/mmol) hypoalbuminaemia oedema

78. Plasma/Serum creatinine

79. GFR

MDRD formula:

$GFR = 170 \times [\text{serum creatinine(mg/dl)}]^{-0.999} \times [\text{age}]^{-0.176} \times [\text{serum urea(mg/dl}]^{-0.17} \times [\text{serum albumin(g/dl)}]^{0.318} \times [0.762 \text{ if female}] \times [1.180 \text{ if black}]$

conversion:

serum creatinine - mg/dl = (μmol/l)/88.5

serum urea - mg/dl = (mmol/l) x 2.8

creatinine clearance not recommended as it is not reliable

80. Active urinary sediment

Uncentrifuged specimen: pyuria (> 5 WCC/hpf), haematuria (> 5 RBC/hpf) or red cell casts in absence of other causes

81. Histology of active nephritis

WHO Class III, IV or V

within last 3 months or since previous assessments if seen less than 3 months ago

glomerular sclerosis without inflammation not counted

OPHTHALMIC

82. Orbital inflammation

83. Severe keratitis

sight threatening

includes: corneal melt

peripheral ulcerative keratitis

84. Mild keratitis

not sight threatening

85. Anterior uveitis

86. Severe posterior uveitis &/or retinal vasculitis

sight-threatening &/or retinal vasculitis

not due to vaso-occlusive disease

87. Mild posterior uveitis &/or retinal vasculitis

not sight-threatening

not due to vaso-occlusive disease

88. Episcleritis

89. Severe scleritis

necrotising anterior scleritis

anterior &/or posterior scleritis requiring systemic steroids/immunosuppression &/or not

	responding to NSAIDs
90. Mild scleritis	anterior &/or posterior scleritis not requiring systemic steroids excludes necrotising anterior scleritis
91. Retinal/choroidal vaso-occlusive disease	includes: retinal arterial & venous occlusion serous retinal &/or retinal pigment epithelial detachments secondary to choroidal vasculopathy
92. Isolated cotton-wool spots	also known as cytoid bodies
93. Optic neuritis	excludes anterior ischaemic optic neuropathy
94. Anterior ischaemic optic neuropathy	visual loss with pale swollen optic disc due to occlusion of posterior ciliary arteries

HAEMATOLOGICAL

95. Haemoglobin	
96. White cell count	
97. Neutrophil count	
98. Lymphocyte count	
99. Platelet count	
100. Evidence of active haemolysis	positive Coomb's test & evidence of haemolysis (raised bilirubin or raised reticulocyte count or reduced haptoglobulins)
101. Isolated positive Coomb's test	

ADDITIONAL ITEMS

These items are required mainly for calculation of GFR

i. Date of Birth
ii. Weight
iii. Black
iv. Serum urea
v. Serum albumin

BILAG2004 INDEX SCORING

- scoring based on the principle of physician's intention to treat

Grade	Definition
A	**Severe disease activity requiring any of the following treatment:** **1. systemic high dose oral corticosteroids (equivalent to prednisolone > 20 mg/day)** **2. intravenous pulse corticosteroids (equivalent to pulse methylprednisolone ≥ 500 mg)** **3. systemic immunosuppressives** **4. therapeutic high dose anticoagulation** **eg: warfarin with target INR 3 - 4**
B	**Moderate disease activity requiring any of the following treatment:** **1. systemic low dose oral corticosteroids (equivalent to prednisolone ≤ 20 mg/day)** **2. intramuscular or intra-articular or soft tissue corticosteroids injection (equivalent to methylprednisolone < 500mg)** **3. topical corticosteroids** **4. topical immunosuppressants** **5. antimalarials** **6. symptomatic therapy** **eg: NSAIDs for inflammatory arthritis** **antipsychotic for psychosis**
C	**Stable mild disease**
D	**Inactive disease but previously affected**
E	**System never involved**

CONSTITUTIONAL

Grade A:
Pyrexia recorded as 2 (same), 3 (worse) or 4 (new) **AND**

Any 2 or more of the following recorded as 2 (same), 3 (worse) or 4 (new):

Weight loss
Lymphadenopathy/splenomegaly
Fatigue/malaise/lethargy
Anorexia

Grade B:
Pyrexia recorded as 2 (same), 3 (worse) or 4 (new) **OR**

Any 2 or more of the following recorded as 2 (same), 3 (worse) or 4 (new):

Weight loss
Lymphadenopathy/splenomegaly
Fatigue/malaise/lethargy
Anorexia

BUT do not fulfil criteria for Grade A

Grade C
Pyrexia recorded as 1 (improving) **OR**

One or more of the following recorded as > 0:

Weight loss
Lymphadenopathy/Splenomegaly
Fatigue/malaise/lethargy
Anorexia

BUT does not fulfil criteria for Grade A or B

Grade D
Previous involvement

Grade E
No previous involvement

MUCOCUTANEOUS

Grade A
Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Skin eruption - severe
Angio-oedema - severe
Mucosal ulceration - severe
Panniculitis/Bullous lupus - severe
Cutaneous vasculitis/thrombosis

Grade B
Any Grade A features recorded as 1 (improving) **OR**

Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Skin eruption - mild
Panniculitis/Bullous lupus - mild
Digital infarcts/nodular vasculitis
Alopecia - severe

Grade C
Any Grade B features recorded as 1 (improving) **OR**

Any of the following recorded as > 0:

Angio-oedema - mild
Mucosal ulceration - mild
Alopecia - mild
Periungual erythema/chilblains
Splinter haemorrhages

Grade D
Previous involvement

Grade E
No previous involvement

NEUROLOGICAL

Grade A
Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Aseptic meningitis
Cerebral vasculitis
Demyelinating syndrome
Myelopathy
Acute confusional state
Psychosis
Acute inflammatory demyelinating polyradiculoneuropathy
Mononeuropathy (single/multiplex)
Cranial neuropathy
Plexopathy
Polyneuropathy
Status epilepticus
Cerebellar ataxia

Grade B
Any Grade A features recorded as 1 (improving) **OR**

Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Seizure disorder
Cerebrovascular disease (not due to vasculitis)
Cognitive dysfunction
Movement disorder
Autonomic disorder
Headache severe unremitting
Headache due to raised intracranial hypertension
Mood disorder

Grade C
Any Grade B features recorded as 1 (improving) **OR**

Any of the following recorded as > 0:

Headache - migraine/cluster/tension

Grade D
Previous involvement

Grade E
No previous involvement

MUSCULOSKELETAL

Grade A
Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Definite Myositis (Bohan and Peter)
Severe polyarthritis

Grade B
Any Grade A features recorded as 1 (improving) **OR**

Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Myositis with incomplete criteria
Arthritis/Tendonitis

Grade C
Any Grade B features recorded as 1 (improving) **OR**

Any of the following recorded as > 0:

Arthralgia/Myalgia

Grade D
Previous involvement

Grade E
No previous involvement

CARDIORESPIRATORY

Grade A
Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Myocarditis/Endocarditis + Cardiac failure
Arrhythmia
New valvular dysfunction
Cardiac tamponade
Pleural effusion with dyspnoea
Pulmonary haemorrhage/vasculitis
Interstitial alveolitis/pneumonitis
Shrinking lung syndrome
Aortitis
Coronary vasculitis

Grade B
Any Grade A features recorded as 1 (improving) **OR**

Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Serositis (pleuro-pericardial pain) - mild
Myocarditis - mild

Grade C
Any Grade B features recorded as 1 (improving)

Grade D
Previous involvement

Grade E
No previous involvement

GASTROINTESTINAL

Grade A
Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Peritonitis
Lupus enteritis/colitis
Intestinal pseudo-obstruction
Acute cholecystitis
Acute pancreatitis

Grade B
Any Grade A feature recorded as 1 (improving) **OR**

Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Abdominal serositis and/or ascites
Malabsorption
Protein losing enteropathy
Hepatitis

Grade C
Any Grade B features recorded as 1 (improving)

Grade D
Previous involvement

Grade E
No previous involvement

OPHTHALMIC

Grade A
Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Orbital inflammation/myositis/proptosis
Keratitis - severe
Posterior uveitis/retinal vasculitis - severe
Scleritis - severe
Retinal/choroidal vaso-occlusive disease
Optic neuritis
Anterior ischaemic optic neuropathy

Grade B
Any Grade A features recorded as 1 (improving) **OR**

Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Keratitis - mild
Anterior uveitis
Posterior uveitis/retinal vasculitis - mild
Scleritis - mild

Grade C
Any Grade B features recorded as 1 (improving) **OR**

Any of the following recorded as > 0:

Episcleritis
Isolated cotton-wool spots (cytoid bodies)

Grade D
Previous involvement

Grade E
No previous involvement

RENAL

Grade A
Two or more of the following **providing 1, 4 or 5 is included**:

1. Deteriorating proteinuria (severe) defined as

   (a) urine dipstick increased by ≥ 2 levels; or
   (b) 24 hour urine protein rising from < 0.20 g to > 1 g; or
   (c) 24 hour urine protein rising from > 1 g by ≥ 100%; or
   (d) newly documented 24 hour urine protein of > 1 g; or
   (e) urine protein-creatinine ratio rising from < 20 mg/mmol to > 100 mg/mmol; or
   (f) urine protein-creatinine ratio rising from > 100 mg/mmol by ≥ 100%; or
   (g) newly documented urine protein-creatinine ratio of > 100 mg/mmol; or
   (h) urine albumin-creatinine ratio rising from < 20 mg/mmol to > 100 mg/mmol ; or
   (i) urine albumin-creatinine ratio rising from > 100 mg/mmol by ≥ 100%; or
   (j) newly documented urine albumin-creatinine ratio of > 100 mg/mmol

2. Accelerated hypertension
3. Deteriorating renal function (severe) defined as

   (a) plasma creatinine > 130 μmol/l and having risen to > 130% of previous value; or
   (b) GFR having fallen to < 67% of previous value; or
   (c) GFR < 50 ml/min, and last time was > 50 ml/min or was not measured.

4. Active urinary sediment
5. Histological evidence of active nephritis within last 3 months
6. Nephrotic syndrome

Grade B
One of the following:

1. One of the Grade A feature

2. Deteriorating proteinuria (mild) defined as

   (a) urine dipstick which has risen by 1 level to at least 2+; **or**
   (b) 24 hour urine protein rising from > 1 g by > 50% but < 100%; **or**
   (c) urine protein-creatinine ratio rising from > 100 mg/mmol by > 50% but < 100%; **or**
   (d) urine albumin-creatinine ratio rising from > 100 mg/mmol by > 50% but < 100%

3. Plasma creatinine > 130 μmol/l and having risen to ≥ 115% but ≤ 130% of previous value

Grade C
One of the following:

1. Mild/Stable proteinuria defined as

(a) urine dipstick ≥ 1+ but has not fulfilled criteria for Grade A & B; **or**
(b) 24 hour urine protein > 0.25 g but has not fulfilled criteria for Grade A & B ; **or**
(c) urine protein-creatinine ratio > 25 mg/mmol but has not fulfilled criteria for Grade A & B; **or**

(d) urine albumin-creatinine ratio > 25 mg/mmol but has not fulfilled criteria for Grade A & B

2. Rising blood pressure (providing the recorded values are > 140/90 mm Hg) which has not fulfilled criteria for Grade A & B, defined as

(a) systolic rise of ≥ 30 mm Hg; **and**
(b) diastolic rise of ≥ 15mm Hg

Grade D
Previous involvement

Grade E
No previous involvement

Note: although albumin-creatinine ratio and protein-creatinine ratio are different, the same cut-off values are used for both in this index

HAEMATOLOGICAL

Grade A
TTP recorded as 2 (same), 3 (worse) or 4 (new) **OR**

Any of the following:

Haemoglobin	< 8 g/dl
Haematocrit	< 24 %
White cell count	< 1.0 x 10^9/l
Neutrophil count	< 0.5 x 10^9/l
Platelet count	< 25 x 10^9/l

Grade B
TTP recorded as 1 (improving) **OR**

Any of the following:

Haemoglobin	8 - 9.9 x 10^9/l
Haematocrit	24 - 29 %
White cell count	1 - 2.4 x 10^9/l
Neutrophil count	0.5 - 1.4 x 10^9/l
Platelet count	25 - 99 x 10^9/l
Evidence of active haemolysis	

Grade C
Any of the following:

Haemoglobin	10 - 10.9 g/dl
Haematocrit	30 - 32 %
White cell count	2.5 - 3.9 x 10^9/l
Neutrophil count	1.5 - 1.9 x 10^9/l
Lymphocyte count	< 1.0 x 10^9/L
Platelet count	100 - 149 x 10^9/l
Isolated Coombs' test positive	

Grade D
Previous involvement

Grade E
No previous involvement

Appendix 2 Classic BILAG Index

All features must be attributable to SLE and refer to last 4 weeks compared with prior disease activity

Indicate features which are present: 1) Improving
2) Same
3) Worse
4) New
Or Y/N or value (where indicated)
(default is 0 = not present)

GENERAL

1. Pyrexia (documented)		( )
2. Weight loss - unintentional > 5%		( )
3. Lymphadenopathy/splenomegaly		( )
4. Fatigue/malaise/lethargy		( )
5. Anorexia/nausea/vomiting		( )

MUCOCUTANEOUS

6. Maculopapular eruption - severe, active (or discoid/bullous)		( )
7. Maculopapular eruption - mild		( )
8. Active discoid lesions – generalised or extensive		( )
9. Active discoid lesions - localised (including lupus profundus)		( )
10. Alopecia (severe, active)		( )
11. Alopecia (mild)		( )
12. Panniculitis (severe)		( )
13. Angio-oedema		( )
14. Extensive mucosal ulceration		( )
15. Small mucosal ulcers		( )
16. Malar erythema		( )
17. Subcutaneous nodules		( )
18. Perniotic skin lesions		( )
19. Peri-ungual erythema		( )
20. Swollen fingers	*Y/N*	( )
21. Sclerodactyly	*Y/N*	( )
22. Calcinosis	*Y/N*	( )
23. Telangiectasia	*Y/N*	( )

NEUROLOGICAL

24. Deteriorating level of consciousness		( )
25. Acute psychosis or delirium or confusional state		( )
26. Seizures		( )
27. Stroke or stroke syndrome		( )
28. Aseptic meningitis		( )
29. Mononeuritis multiplex		( )
30. Ascending or transverse myelitis		( )
31. Peripheral or cranial neuropathy		( )
32. Disc swelling/cytoid bodies		( )

33. Chorea		( )
34. Cerebellar ataxia		( )
35. Headache severe, unremitting		( )
36. Organic depressive illness		( )
37. Organic brain syndrome including pseudotumor cerebri		( )
38. Episodic migraine headaches		( )

MUSCULOSKELETAL

39. Definite myositis (Bohan & Peter)		( )
40. Severe polyarthritis - with loss of function		( )
41. Arthritis		( )
42. Tendonitis		( )
43. Mild chronic myositis		( )
44. Arthralgia		( )
45. Myalgia		( )
46. Tendon contractures and fixed deformity	*Y/N*	( )
47. Aseptic necrosis	*Y/N*	( )

CARDIOVASCULAR & RESPIRATORY

48. Pleuropericardial pain		( )
49. Dyspnoea		( )
50. Cardiac failure		( )
51. Friction rub		( )
52. Effusion (pericardial or pleural)		( )
53. Mild or intermittent chest pain		( )
54. Progressive Chest X-ray changes-lung fields	*Y/N*	( )
55. Progressive Chest X-ray changes-heart size	*Y/N*	( )
56. ECG evidence of pericarditis or myocarditis	*Y/N*	( )
57. Cardiac arrhythmias including tachycardia> 100 in absence of fever	*Y/N*	( )
58. Pulmonary function fall by > 20%	*Y/N*	
59. Cytohistological evidence of inflammatory lung disease	*Y/N*	( )

VASCULITIS

60. Major cutaneous vasculitis including ulcers		( )
61. Major abdominal crisis due to vasculitis		( )
62. Recurrent thromboembolism (excluding strokes)		( )
63. Raynaud's		( )
64. Livedo reticularis		( )
65. Superficial phlebitis		( )
66. Minor cutaneous vasculitis(nail fold vasculitis, digital vasculitis, purpura, urticaria)		( )
67. Thromboembolism (excluding stroke) - 1st episode	*Y/N*	( )

RENAL

68. Systolic blood pressure (mm Hg)	*value*	( )
69. Diastolic blood pressure (mm Hg)	*value*	( )
70. Accelerated hypertension	*Y/N*	( )

71. Urine dipstick protein (+ = 1, ++ = 2, +++ = 3)	*value*	( )
72. 24 hour urinary protein (g)	*value*	( )
73. Newly documented proteinuria > 1g / 24hrs	*Y/N*	( )
74. Nephrotic syndrome	*Y/N*	( )
75. Creatinine (plasma/serum)	*value*	( )
76. Creatinine clearance/GFR ml/min	*value*	( )
77. Active urinary sediment	*Y/N*	( )
78. Histological evidence of active nephritis - within 3 months	*Y/N*	( )

HAEMATOLOGY

79. Haemoglobin g/dl	*value*	( )
80. Total white cell count x 109/l	*value*	( )
81. Neutrophils x 109/L	*value*	( )
82. Lymphocytes x 109/L	*value*	( )
83. Platelets x 109/L	*value*	( )
84. Evidence of active haemolysis	*Y/N*	( )
85. Coomb's test positive	*Y/N*	( )
86. Evidence of circulating anticoagulant	*Y/N*	( )

CLASSIC BILAG GLOSSARY

It is implicit in this scoring system that all features scored are considered to be due to active lupus.

GENERAL

1. Pyrexia: temperature > 37.5°C documented
2. Weight loss unintentional more than 5%
3. Lymphadenopathy: palpable lymph nodes > 1cm in diameter
4. Fatigue or malaise or lethargy
5. Anorexia or nausea or vomiting

MUCOCUTANEOUS

6. Maculopapular eruption – severe: active maculopapular or bullous eruption; extensive (> 18% of body surface area), scarring or causing disability.
7. Maculopapular eruption – mild (≤ 18% body surface area), non-scarring, non-disabling
8. Active discoid lesions – generalised or extensive, > 18% surface area
9. Active discoid lesions – localised ie ≤ 18% surface area
10. Alopecia – severe, active: abnormal diffuse hair loss which is clinically detectable with scalp inflammation.
11. Alopecia (mild): limited relatively inactive: abnormal diffuse hair loss with little or no detectable scalp inflammation.
12. Panniculitis: extensive, painful, erythematous subcutaneous nodules associated with fat necrosis resolve with scarring, or lupus profundus: erythematous elevated plaques with an overlying discoid skin lesion
13. Angio-oedema: potentially life threatening eg stridor
14. Extensive mucosal ulceration: severe, deep, disabling ulcers.

NEUROLOGICAL

24. Acute deteriorating level of consciousness by any accepted clinical criteria
25. Acute psychosis or delirium or confusional state: severe disturbance in the perception of reality characterised by: delusions, hallucinations, incoherence, marked illogical thinking, bizarre or catatonic behaviour
35. Headache severe, unremitting: continuous headache not relieved by narcotic analgesia
36. Organic depressive illness: associated with somatic symptoms and severe enough to merit treatment with anti-depressive medication
37. Organic brain syndrome: impaired orientation, memory or other intellectual function in the absence of metabolic, psychiatric or pharmacological causes
    Clinical features develop over a short period (usually hours to days) and tend to fluctuate over the course of the day:
    a. Clouding of consciousness with reduced capacity to focus and sustain attention to environment.
    b. i. Perceptual disturbance: misinterpretations, illusions or hallucinations
       ii. incoherent speech
       iii. insomnia or daytime drowsiness

iv. increased or decreased psychomotor activity

c. Disorientation and recent memory impairment

MUSCULOSKELETAL

39. Myositis: at least three of – proximal muscle weakness, elevated muscle enzymes, positive muscle biopsy and abnormal EMG.
40. Polyarthritis with loss of function: active joint inflammation with clinically significant loss of the functional range of movement of the involved joints.
41. Arthritis: active joint inflammation without loss of functional range of motion

CARDIO RESPIRATORY

48. Pleuropericardial pain: localised sharp or dull pain in the chest aggravated by respiration.
49. Dyspnoea: on exercise (not orthopnoea alone).
53. Mild intermittent chest pain: non-specific (not clearly pleuritic, pericardial, musculoskeletal or angina).
58. Pulmonary function fall by > 20%: >20% of expected (predicted for height, weight, sex and age) or > 20% fall in total lung capacity (forced vital capacity) and/or DLCO.

VASCULAR SYSTEM

59. Major cutaneous vasculitis including ulcers: extensive gangrene and/or ulceration
66. Minor cutaneous vasculitis eg digital vasculitis with nail fold infarcts

RENAL

70. Accelerated hypertension: BP rising to >170/110 (5^{th} phase) within one month, if accompanied by Grade IV retinal changes (ie haemorrhage, exudates).
77. Active urinary sediment: on uncentrifuged specimen. Pyuria (> 5wc/hpf), haematuria (>5rbc/hpf) or red cell casts in the absence of infection or any other cause
78. Histological evidence of active nephritis: according to WHO criteria. Sclerosis alone (without inflammation) will not be regarded as active nephritis

HAEMATOLOGY

79. Evidence of circulating lupus anticoagulant or other antiphospholipid antibody.

CLASSIC BILAG INDEX SCORING

General

Category A
Pyrexia scoring 2 (same), 3 (worse) or 4 (new) **AND**
≥ 2 others scoring 2 (same), 3 (worse) or 4 (new)

Category B
Pyrexia scoring 2 (same), 3 (worse) or 4 (new) **OR**
≥ 2 others scoring 2 (same), 3 (worse) or 4 (new)

BUT do not fulfill criteria for Category A

Category C
Pyrexia scoring 1 (improving) **OR**
One or more of others scoring > 0

BUT do not fulfill criteria for Category A or B

Category D
Previous involvement

Category E
No previous involvement

Mucocutaneous

Category A
Any one of the following scoring 2 (same), 3 (worse) or 4 (new):

Maculopapular/Bullous eruption - severe
Discoid lesions – extensive
Angio-oedema
Extensive mucosal ulceration

Category B
Any one of the following scoring 2 (same), 3 (worse) or 4 (new)

Malar erythema
Maculopapular eruption - mild
Panniculitis
Discoid lesions - localised
Alopecia - severe
Subcutaneous nodules
Perniotic skin lesions

Category C
Any Category A or Category B criteria scoring 1 (improving) **OR**

Any one of the following scoring > 0:

Peri-ungual erythema
Swollen fingers
Sclerodactyly
Calcinosis
Telangiectasia
Alopecia - mild
Mucosal ulceration -small

Category D
Previous involvement

Category E
No previous involvement

Neurological

Category A
Any one of the following scoring 3 (worse) or 4 (new):

Deteriorating level of consciousness
Acute psychosis or delirium or confusional state
Seizures
Stroke or stroke syndrome
Aseptic meningitis
Mononeuritis multiplex
Ascending or transverse myelitis
Peripheral or cranial neuropathy
Chorea
Cerebellar ataxia

Category B
Any one of the following scoring 3 (worse) or 4 (new):

Headache (severe unremitting)
Organic depressive illness
Organic brain syndrome including pseudotumor cerebri
Disc swelling or cytoid bodies

OR

Any of the following scoring 2 (same) or 1 (improving):

Deteriorating level of consciousness
Acute psychosis or delirium or confusional state
Seizure

Category C

Episodic migraine headaches scoring > 0 OR

Any one of the following scoring 2 (same) or 1 (improving)

Stroke or stroke syndrome
Aseptic meningitis
Mononeuritis multiplex
Ascending or transverse myelitis
Peripheral or cranial neuropathy
Chorea
Cerebellar ataxia
Headache (severe unremitting)
Organic depressive illness
Organic brain syndrome including pseudotumor cerebri

Disc swelling or cytoid bodies

Category D
Previous involvement

Category E
No previous involvement

Musculoskeletal

Category A
One or more of the following scoring > 1:

Definite myositis (Bohan and Peter)
Severe polyarthritis

Category B
One or more of the following scoring > 1:
Arthritis (definite synovitis)
Tendonitis

Category C
Any Category A or Category B criteria scoring 1 **OR**

One or more of the following scoring > 0 (or Yes)

Arthralgia
Myalgia
Tendon contractures and fixed deformity
Aseptic necrosis
Mild chronic myositis

Category D
Previous involvement

Category E
No previous involvement

Cardiorespiratory

Category A
Cardiac failure scoring >1 plus ≥ 2 other criteria listed below scoring >1

OR

Symptomatic effusion scoring > 1 plus ≥ 2 other criteria listed below scoring >1

OR

Four of the criteria listed below each scoring > 1

Pleuropericardial pain
Dyspnoea
Friction Rub
Progressive Chest X-ray changes – lung fields
Progressive Chest X-ray changes – heart size
ECG evidence of pericarditis or myocarditis
Cardiac arrhythmias including tachycardia - >100 in absence of fever
Deteriorating lung function: <20% of expected or >20% fall
Cytohistological evidence of inflammatory lung disease

Category B
Two of the criteria listed below each scoring > 1:

Pleuropericardial pain
Dyspnoea
Friction Rub
Progressive Chest X-ray changes – lung fields
Progressive Chest X-ray changes – heart size
ECG evidence of pericarditis or myocarditis
Cardiac arrhythmias including tachycardia - >100 in absence of fever
Deteriorating lung function: <20% of expected or >20% fall
Cytohistological evidence of inflammatory lung disease

Category C
One of the criteria listed below each scoring > 0:

Mild intermittent chest pain
Pleuropericardial pain
Dyspnoea
Friction Rub
Progressive Chest X-ray changes – lung fields
Progressive Chest X-ray changes – heart size
ECG evidence of pericarditis or myocarditis
Cardiac arrhythmias including tachycardia - >100 in absence of fever
Deteriorating lung function: <20% of expected or >20% fall

Cytohistological evidence of inflammatory lung disease

Category D
Previous involvement

Category E
No previous involvement

Vasculitis

Category A
Any of the following scoring > 1:

Major cutaneous vasculitis
Major abdominal crisis due to vasculitis
Recurrent thromboembolism (excluding strokes)

Category B
Any of the following scoring > 1:

Minor cutaneous vasculitis
Superficial phlebitis
Thromboembolism (excluding strokes), first episode

Category C

Any Category A or Category B criteria scoring 1
Raynaud's phenomenon, scoring > 0
Livedo reticularis, scoring > 0

Category D
Previous involvement

Category E
No previous involvement

Renal

Category A
Two or more of the following providing **1, 4 or 5 is included**:
1. Proteinuria, defined as
    (a) urinary dipstick increased by 2 or more levels; **or**
    (b) 24 urinary protein rising from < 0.20g to > 1g; **or**
    (c) 24 hour urinary protein rising from > 1g by 100%; **or**
    (d) newly documented proteinuria of > 1g

2. Accelerated hypertension
3. Deteriorating renal function, defined as
    (a) plasma creatinine >130 μmol/l and having risen to >130% of previous value; **or**
    (b) creatinine clearance having fallen to < 67% of previous value; **or**
    (c) creatinine clearance < 50 ml/min, and last time was > 50 ml/min or was not measured

4. Active urinary sediment
5. Histological evidence of active nephritis

Category B
One of the following:
1. Any one of the Category A criteria above
2. Proteinuria
    (a) urinary dipstick which has risen by 1+ or more to at least 2+, **or**
    (b) 24 hour urinary protein rising from > 1g by >50% but <100%

3. Plasma creatinine > 130 μmol/l and having risen to 115% of previous value

Category C
One of the following:

24 hour urinary protein > 0.25g
Urinary dipstick 1+ or more
Rising blood pressure, defined as
    (i) systolic rise of > 30mm **and**
    (ii) diastolic rise of >15mm **and**
    (iii) the recorded values are > 140/90)

Category D
Previous involvement

Category E
No previous involvement

Haematological

Category A
One of the following:

White cell count < 1000
Platelet count < 25
Haemoglobin < 8

Category B
One of the following:

White cell count < 2500
Platelet count < 100
Haemoglobin < 11
Coombs Test positive **and** evidence of active haemolysis, e.g.
raised bilirubin +/- increased Reticulocyte Count

Category C
One of the following:

White cell count < 4000
Platelet count < 150
Lymphocyte count < 1500
Coombs Test positive but no evidence of active haemolysis
Evidence of circulating lupus anticoagulant or other antiphospholipid antibody

Category D
Previous involvement

Category E
No previous involvement

Appendix 3 SLEDAI-2000 Index

(**Circle in SLEDAI Score column** if descriptor is present at the time of the visit or in the **preceding 10 days**)

Item	SLEDAI SCORE	Descriptor	Definition
1	8	Seizure	Recent onset, exclude metabolic, infectious or drug causes
2	8	Psychosis	Altered ability to function in normal activity due to severe disturbance in the perception of reality. Include hallucinations, incoherence, marked loose associations, impoverished thought content, marked illogical thinking, bizarre, disorganised, or catatonic behaviour. Exclude uraemia and drug causes
3	8	Organic brain syndrome	Altered mental function with impaired orientation, memory, or other intellectual function, with rapid onset and fluctuating clinical features, inability to sustain attention to environment, plus at least 2 of the following: perceptual disturbance, incoherent speech, insomnia or daytime drowsiness, or increased or decreased psychomotor activity. Exclude metabolic, infectious or drug causes
4	8	Visual disturbance	Retinal changes of SLE. Include cytoid bodies, retinal hemorrhages, serous exudates or hemorrhages in the choroid, or optic neuritis. Exclude hypertension, infection, or drug causes
5	8	Cranial nerve disorder	New onset of sensory or motor neuropathy involving cranial nerves
6	8	Lupus headache	Severe, persistent headache; may be migrainous, but must be non-responsive to narcotic analgesia
7	8	CVA	New onset Cerebrovascular accident(s). Exclude arteriosclerosis
8	8	Vasculitis	Ulceration, gangrene, tender finger nodules, periungual infarction, splinter hemorrhages or biopsy or angiogram proof of vasculitis
9	4	Arthritis	≥ 2 joints with pain and signs of inflammation (i.e. tenderness, swelling or effusion)
10	4	Myositis	Proximal muscle aching/weakness, associated with elevated creatinine phosphokinase (CK)/aldolase, or EMG changes or a biopsy showing myositis
11	4	Urinary casts	Heme-granular or RBC casts
12	4	Hematuria	> 5 RBC/high power field. Exclude stone, infection or other cause
13	4	Proteinuria	> 0.5 gram/24 hours
14	4	Pyuria	> 5 WBC/high power field. Exclude infection
15	2	Rash	Inflammatory type rash
16	2	Alopecia	Abnormal, patchy or diffuse loss of hair
17	2	Mucosal ulcers	Oral or nasal ulcerations
18	2	Pleurisy	Pleuritic chest pain with pleural rub or effusion, or pleural thickening
19	2	Pericarditis	Pericardial pain with at least 1 of the following: rub, effusion or ECG or echocardiogram confirmation
20	2	Low complement	Decrease in CH50, C3 or C4 below lower limit of normal for testing laboratory
21	2	Increased DNA binding	Increased DNA binding above normal range for testing laboratory
22	1	Fever	> 38ºC. Exclude infectious cause

23	1	Thrombocyto penia	< 100 x 10^9 platelets/L, exclude drug causes
24	1	Leukopenia	< 3 x 10^9 WBC/L, exclude drug causes

TOTAL SCORE:

Appendix 4 SLICC/ACR Damage Index

Damage (non-reversible change, not related to active inflammation) occurring since diagnosis of lupus, ascertained by clinical assessment and present for at least 6 months unless otherwise stated. Repeat episodes must occur at least 6 months apart to score 2. The same lesion cannot be scored twice.

	Score	
OCULAR		
Any cataract ever (documented by ophthalmoscopy)	**1**	
Retinal change ***OR*** Optic atrophy (documented by ophthalmoscopy)	**1**	
NEUROPSYCHIATRIC		
Cognitive impairment (eg memory deficit, difficulty with calculation, poor concentration, difficulty in spoken or written language, impaired performance level) ***OR*** Major psychosis	**1**	
Seizures requiring therapy for 6 months	**1**	
Cerebrovascular accident or surgical resection (for non-malignant causes) ***(score 2 if >1)***	**1**	**2**
Cranial or peripheral neuropathy (excluding optic)	**1**	
Transverse myelitis	**1**	
RENAL		
Estimated/Measured GFR < 50%	**1**	
Proteinuria ≥ 3.5g/24 hours ***OR***	**1**	
End-stage renal failure (regardless of dialysis or transplantation)	**3**	
PULMONARY		
Pulmonary hypertension (right ventricular prominence or loud P2)	**1**	
Pulmonary fibrosis (physical & radiograph)	**1**	
Shrinking lung (radiograph)	**1**	
Pleural fibrosis (radiograph)	**1**	
Pulmonary infarction (radiograph) or resection (for non-malignant causes)	**1**	
CARDIOVASCULAR		
Angina ***OR*** Coronary artery bypass	**1**	
Myocardial infarction ***(score 2 if > 1)***	**1**	**2**
Cardiomyopathy (ventricular dysfunction)	**1**	
Valvular disease (diastolic, murmur, or systolic murmur > 3/6)	**1**	
Pericarditis for 6 months ***OR*** Pericardiectomy	**1**	
PERIPHERAL VASCULAR		
Claudication for 6 months	**1**	
Minor tissue loss (pulp space)	**1**	
Significant tissue loss (eg loss of digit or limb) ***(score 2 if > 1)***	**1**	**2**
Venous thrombosis with swelling ulceration ***OR*** Venous stasis	**1**	
GASTROINTESTINAL		
Infarction or resection of bowel below duodenum, spleen, liver or gallbladder for any cause ***(score 2 if > 1)***	**1**	**2**
Mesenteric insufficiency	**1**	
Chronic peritonitis	**1**	
Stricture OR Upper gastrointestinal surgery	**1**	
Pancreatic insufficiency requiring enzyme replacement	**1**	
MUSCULOSKELETAL		

Muscle atrophy or weakness	**1**	
Deforming or erosive arthritis (including reversible deformities, excluding avascular necrosis)	**1**	
Osteoporosis with fracture or vertebral collapse (excluding avascular necrosis)	**1**	
Avascular necrosis (imaging) ***(score 2 if > 1)***	**1**	**2**
Osteomyelitis (supported by culture evidence)	**1**	
Tendon rupture	**1**	
SKIN		
Scarring chronic alopecia	**1**	
Extensive scarring or panniculum other than scalp and pulp space	**1**	
Skin ulceration for > 6 months (excluding thrombosis)	**1**	
PREMATURE GONADAL FAILURE (secondary amenorrhoea before age 40)	**1**	
DIABETES MELLITUS (regardless of treatment)	**1**	
MALIGNANCY (exclude dysplasia) ***(score 2 if > 1 site)***	**1**	**2**

Appendix 5 Agreement Tables from Reliability Study

1) Constitutional

Exercise 1

		Rater 2				Total
		A	**B**	**C**	**D**	
Rater 1	**A**	0	0	0	1	1
	B	0	3	1	7	11
	C	0	2	15	34	51
	D	0	1	4	29	34
Total		0	6	20	71	97

	Agreement (%)	Expected Agreement (%)	Kappa
Unweighted	48.45	37.20	0.18
Weighted	69.59	62.95	0.18

Exercise 2

		Rater 2				Total
		A	**B**	**C**	**D**	
Rater 1	**A**	1	0	0	0	1
	B	0	3	2	3	8
	C	0	4	13	27	44
	D	0	0	3	25	28
Total		1	7	18	55	81

	Agreement (%)	Expected Agreement (%)	Kappa
Unweighted	53.61	38.72	0.24
Weighted	75.26	64.08	0.31

2) Mucocutaneous

Exercise 1

		Rater 2				Total
		A	**B**	**C**	**D**	
Rater 1	**A**	2	0	2	1	5
	B	1	11	2	5	19
	C	1	4	11	9	25
	D	0	1	4	43	48
Total		4	16	19	58	97

	Agreement (%)	Expected Agreement (%)	Kappa
Unweighted	69.07	38.08	0.50
Weighted	80.41	56.44	0.55

Exercise 2

		Rater 2				Total
		A	**B**	**C**	**D**	
Rater 1	**A**	0	1	0	0	1
	B	0	13	3	2	18
	C	0	6	22	4	32
	D	0	3	1	26	30
Total		0	23	26	32	81

	Agreement (%)	Expected Agreement (%)	Kappa
Unweighted	72.16	32.96	0.58
Weighted	84.28	55.36	0.65

3) Neuropsychiatric

Exercise 1

		Rater 2				Total
		A	**B**	**C**	**D**	
Rater 1	**A**	3	0	0	5	8
	B	0	0	0	1	1
	C	0	0	8	16	24
	D	0	0	3	61	64
Total		3	0	11	83	97

	Agreement (%)	**Expected Agreement (%)**	**Kappa**
Unweighted	74.23	59.52	**0.36**
Weighted	84.02	74.35	**0.38**

Exercise 2

		Rater 2				Total
		A	**B**	**C**	**D**	
Rater 1	**A**	2	1	0	0	3
	B	0	2	0	2	4
	C	0	0	5	8	13
	D	1	1	0	59	61
Total		3	4	5	69	81

	Agreement (%)	**Expected Agreement (%)**	**Kappa**
Unweighted	79.38	62.96	0.44
Weighted	85.82	73.09	0.47

4) Musculoskeletal

Exercise 1

		Rater 2				Total
		A	**B**	**C**	**D**	
Rater 1	**A**	1	3	0	0	4
	B	0	5	4	4	13
	C	0	7	9	25	41
	D	0	4	0	35	39
Total		1	19	13	64	97

	Agreement (%)	**Expected Agreement (%)**	**Kappa**
Unweighted	**51.55**	34.86	**0.26**
Weighted	72.42	57.49	**0.35**

Exercise 2

		Rater 2				Total
		A	**B**	**C**	**D**	
Rater 1	**A**	2	0	1	0	3
	B	2	5	4	1	12
	C	0	5	14	16	35
	D	0	2	3	25	30
Total		4	12	22	42	80

	Agreement (%)	**Expected Agreement (%)**	**Kappa**
Unweighted	56.70	34.25	0.34
Weighted	77.32	57.00	0.47

5) Cardiorespiratory

Exercise 1

		Rater 2				Total
		A	**B**	**C**	**D**	
Rater 1	**A**	0	0	0	3	3
	B	1	5	0	2	8
	C	0	1	0	0	1
	D	0	2	0	83	85
Total		1	8	0	88	97

	Agreement (%)	Expected Agreement (%)	Kappa
Unweighted	90.72	80.21	0.53
Weighted	92.01	80.98	0.58

Exercise 2

		Rater 2				Total
		A	**B**	**C**	**D**	
Rater 1	**A**	0	0	0	0	0
	B	0	3	0	5	8
	C	0	0	1	1	2
	D	0	0	1	70	71
Total		0	3	2	76	81

	Agreement (%)	Expected Agreement (%)	Kappa
Unweighted	90.72	83.46	0.44
Weighted	92.27	85.98	0.45

6) Gastrointestinal

Exercise 1

		Rater 2				Total
		A	**B**	**C**	**D**	
Rater 1	**A**	0	0	0	0	0
	B	0	0	0	1	1
	C	0	0	0	1	1
	D	0	1	0	94	95
Total		0	1	0	96	97

	Agreement (%)	Expected Agreement (%)	Kappa
Unweighted	96.91	96.94	-0.01
Weighted	97.42	97.45	-0.01

Exercise 2

		Rater 2				Total
		A	**B**	**C**	**D**	
Rater 1	**A**	0	0	0	0	0
	B	0	0	0	1	1
	C	0	0	0	0	0
	D	0	0	0	80	80
Total		0	0	0	81	81

	Agreement (%)	Expected Agreement (%)	Kappa
Unweighted	98.97	98.97	0
Weighted	98.97	98.97	0

7) Ophthalmic

Exercise 1

		Rater 2				Total
		A	**B**	**C**	**D**	
Rater 1	**A**	0	0	0	0	0
	B	0	0	0	2	2
	C	0	0	0	2	2
	D	0	0	0	93	93
Total		0	0	0	97	97

	Agreement (%)	Expected Agreement (%)	Kappa
Unweighted	95.88	95.88	0
Weighted	96.91	96.91	0

Exercise 2

		Rater 2				Total
		A	**B**	**C**	**D**	
Rater 1	**A**	0	0	0	0	0
	B	0	0	0	1	1
	C	0	0	1	0	1
	D	0	0	0	79	79
Total		0	0	1	80	81

	Agreement (%)	Expected Agreement (%)	Kappa
Unweighted	98.97	96.94	0.66
Weighted	98.97	97.96	0.49

8) Renal

Exercise 1

		Rater 2				Total
		A	**B**	**C**	**D**	
Rater 1	**A**	1	0	0	0	1
	B	0	7	0	0	7
	C	0	0	11	2	13
	D	0	0	1	75	76
Total		1	7	12	77	97

	Agreement (%)	Expected Agreement (%)	Kappa
Unweighted	96.91	64.39	0.91
Weighted	98.45	75.66	0.94

Exercise 2

		Rater 2				Total
		A	**B**	**C**	**D**	
Rater 1	**A**	4	0	0	0	4
	B	0	4	0	0	4
	C	0	0	1	4	5
	D	0	0	0	84	84
Total		4	4	1	88	97

	Agreement (%)	Expected Agreement (%)	Kappa
Unweighted	95.88	78.96	0.80
Weighted	97.94	82.19	0.88

9) Haematological

Exercise 1

		Rater 2				Total
		A	**B**	**C**	**D**	
Rater 1	**A**	0	0	0	0	0
	B	0	2	1	0	3
	C	0	0	43	3	46
	D	0	0	4	41	45
Total		0	2	48	44	94

	Agreement (%)	Expected Agreement (%)	Kappa
Unweighted	91.49	47.46	0.84
Weighted	95.74	72.48	0.85

Exercise 2

		Rater 2				Total
		A	**B**	**C**	**D**	
Rater 1	**A**	0	0	0	0	0
	B	0	8	2	0	10
	C	0	0	23	1	24
	D	0	0	8	55	63
Total		0	8	33	56	97

	Agreement (%)	Expected Agreement (%)	Kappa
Unweighted	88.66	46.76	0.79
Weighted	94.33	67.73	0.82

Appendix 6 Revised BILAG-2004 Index (Post-Reliability Study)

BILAG-2004 INDEX

Only record items due to SLE Disease Activity & assessment refers to manifestations occurring in the last 4 weeks (compared with the previous 4 weeks).

♦♦ TO BE USED WITH THE GLOSSARY ♦♦

Scoring:
ND Not Done
1 Improving
2 Same
3 Worse
4 New
Yes/No OR Value (where indicated)
❑ indicate if not due to SLE activity
(default is 0 = not present)

CONSTITUTIONAL

1. Pyrexia - documented > 37.5ºC	(	)
2. Weight loss - unintentional > 5%	(	)
3. Lymphadenopathy/splenomegaly	(	)
4. Anorexia	(	)

MUCOCUTANEOUS

5. Skin eruption - severe	(	)
6. Skin eruption - mild	(	)
7. Angio-oedema - severe	(	)
8. Angio-oedema - mild	(	)
9. Mucosal ulceration - severe	(	)
10. Mucosal ulceration - mild	(	)
11. Panniculitis/Bullous lupus - severe	(	)
12. Panniculitis/Bullous lupus - mild	(	)
13. Cutaneous vasculitis/thrombosis	(	)
14. Digital infarcts/nodular vasculitis	(	)
15. Alopecia - severe	(	)
16. Alopecia - mild	(	)
17. Peri-ungual erythema/chilblains	(	)
18. Splinter haemorrhages	(	)

NEUROPSYCHIATRIC

19. Aseptic meningitis	(	)
20. Cerebral vasculitis	(	)
21. Demyelinating syndrome	(	)
22. Myelopathy	(	)
23. Acute confusional state	(	)
24. Psychosis	(	)
25. Acute inflammatory demyelinating polyradiculoneuropathy	(	)
26. Mononeuropathy (single/multiplex)	(	)
27. Cranial neuropathy	(	)

28. Plexopathy ( )
29. Polyneuropathy ( )
30. Seizure disorder ( )
31. Status epilepticus ( )
32. Cerebrovascular disease (not due to vasculitis) ( )
33. Cognitive dysfunction ( )
34. Movement disorder ( )
35. Autonomic disorder ( )
36. Cerebellar ataxia (isolated) ( )
37. Lupus headache - severe unremitting ( )
38. Headache from IC hypertension ( )

MUSCULOSKELETAL
39. Definite myositis (Bohan & Peter) ( )
40. Myositis with incomplete criteria ( )
41. Arthritis(severe) ( )
42. Arthritis(moderate)/Tendonitis/Tenosynovitis ( )
43. Arthritis(mild)/Arthralgia/Myalgia ( )

CARDIORESPIRATORY
44. Myocarditis - mild ( )
45. Myocarditis/Endocarditis + Cardiac failure ( )
46. Arrhythmia ( )
47. New valvular dysfunction ( )
48. Serositis (pleuro-pericardial pain) - mild ( )
49. Cardiac tamponade ( )
50. Pleural effusion with dyspnoea ( )
51. Pulmonary haemorrhage/vasculitis ( )
52. Interstitial alveolitis/pneumonitis ( )
53. Shrinking lung syndrome ( )
54. Aortitis ( )
55. Coronary vasculitis ( )

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 ( )

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		(	)	

RENAL

78. Systolic blood pressure (mm Hg)	value	(	)	❑
79. Diastolic blood pressure (mm Hg)	value	(	)	❑
80. Accelerated hypertension	Yes/No	(	)	
81. Urine dipstick protein	(+=1, ++=2, +++=3)	(	)	❑
82. Urine albumin-creatinine ratio	mg/mmol	(	)	❑
83. Urine protein-creatinine ratio	mg/mmol	(	)	❑
84. 24 hour urine protein (g)	value	(	)	❑
85. Nephrotic syndrome	Yes/No	(	)	
86. Creatinine (plasma/serum)	µmol/l	(	)	❑
87. GFR (calculated)	ml/min/1.73 m^2	(	)	❑
88. Active urinary sediment	Yes/No	(	)	
89. Active nephritis	Yes/No	(	)	

HAEMATOLOGICAL

90. Haemoglobin (g/dl)	value	(	)	❑
91. Total white cell count (x 10^9/l)	value	(	)	❑
92. Neutrophils (x 10^9/l)	value	(	)	❑
93. Lymphocytes (x 10^9/l)	value	(	)	❑
94. Platelets (x 10^9/l)	value	(	)	❑
95. TTP		(	)	
96. Evidence of active haemolysis	Yes/No	(	)	
97. Coombs' test positive (isolated)	Yes/No	(	)	

BILAG-2004 INDEX GLOSSARY

INSTRUCTIONS

- only record features that are **attributable to SLE disease activity and not due to damage, infection, thrombosis (in absence of inflammatory process) or other conditions**

- assessment refers to manifestations occurring in the **last 4 weeks compared with the previous 4 weeks**

- activity refers to disease process which is reversible while damage refers to permanent process/scarring (irreversible)

- damage due to SLE should be considered as a cause of features that are fixed/persistent (SLICC/ACR damage index uses persistence ≥ 6 months to define damage)

- in some manifestations, it may be difficult to differentiate SLE from other conditions as there may not be any specific test and the decision would then lies with the **physician's judgement on the balance of probabilities**

- ophthalmic manifestations would usually need to be assessed by ophthalmologist (using the proforma included with this glossary) and these items would need to be recorded retrospectively after receiving the response from ophthalmologist

- guidance for scoring:

(4) NEW

- manifestations are recorded as new when it is a new episode occurring in the last 4 weeks (compared to the previous 4 weeks) which has not improved and this includes new episodes (recurrence) of old manifestations

- new episode occurring in the last 4 weeks but also satisfying the criteria for improvement (below) would be classified as improving instead of new

(3) WORSE

- this refers to manifestations that have deteriorated in the last 4 weeks compared to the previous 4 weeks

(2) SAME

- this refers to manifestations that have been present for the last 4 weeks and the previous 4 weeks without significant improvement or deterioration (from the previous 4 weeks)

- this also applies to manifestations that have improved over the last 4 weeks compared to the previous 4 weeks but do not meet the criteria for improvement

(1) IMPROVING

- definition of **improvement**: (a) the amount of improvement is sufficient for **consideration of reduction in therapy** and would not justify escalation in therapy

  (b) improvement must be **present currently and ≥ 2 weeks** of the last 4 weeks

(0) NOT PRESENT

(ND) NOT DONE

- it is important to indicate if a test has not been performed (particularly laboratory investigations) so that this will be recorded as such in the database & not as normal or absent (which is the default)

❑ **INDICATE (TICK) IF NOT DUE TO SLE ACTIVITY**

- for descriptors that are based on measurements (in renal and haematology systems), it is important to indicate if these are not due to lupus disease activity (for consideration of scoring) as they are usually recorded routinely into a database

CONSTITUTIONAL

1. Pyrexia — temperature > 37.5˚C documented

2. Unintentional weight loss > 5%

3. Lymphadenopathy — palpable lymph node more than 1 cm diameter

exclude infection

4. Anorexia

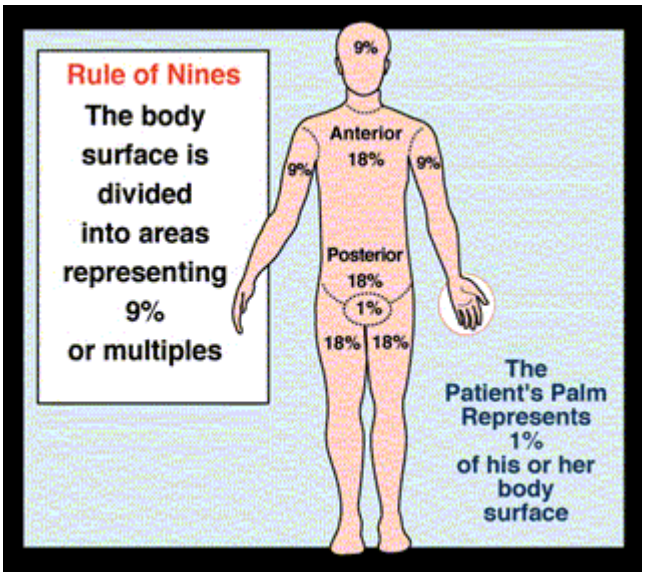
MUCOCUTANEOUS

5. Severe eruption — > 18% body surface area

any lupus rash except panniculitis, bullous lesion & angio-oedema

body surface area (BSA) is estimated using the rules of nines (used to assess extent of burns) as follows:

palm(excluding fingers) = 1% BSA
each lower limb = 18% BSA
each upper limb = 9% BSA
torso (front) = 18% BSA
torso (back) = 18% BSA
head = 9% BSA
genital (male) = 1% BSA



6. Mild eruption — ≤ 18% body surface area

	any lupus rash except panniculitis, bullous lesion & angio-oedema
7. Severe angio-oedema	potentially life-threatening eg: stridor angio-oedema is a variant form of urticaria which affects the subcutaneous, submucosal and deep dermal tissues
8. Mild angio-oedema	not life threatening
9. Severe mucosal ulceration	disabling (significantly interfering with oral intake), extensive & deep ulceration must have been observed by a physician
10. Mild mucosal ulceration	localised &/or non-disabling ulceration
11. Severe panniculitis or bullous lupus	any one: > 9% body surface area facial panniculitis panniculitis that is beginning to ulcerate panniculitis that threatens integrity of subcutaneous tissue (beginning to cause surface depression) on > 9% body surface area panniculitis presents as a palpable and tender subcutaneous induration/nodule note that established surface depression and atrophy alone is likely to be due to damage
12. Mild panniculitis or bullous lupus	≤ 9% body surface area does not fulfill any criteria for severe panniculitis (for panniculitis)
13. Cutaneous vasculitis/thrombosis	resulting in extensive gangrene or ulceration or skin infarction
14. Digital infarct/nodular vasculitis	localised single or multiple infarct(s) over digit(s) or tender erythematous nodule(s)
15. Severe alopecia	clinically detectable (diffuse or patchy) hair loss with scalp inflammation (redness over scalp)

16. Mild alopecia	diffuse or patchy hair loss without scalp inflammation (clinically detectable or by history)
17. Peri-ungual erythema or chilblains	chilblains are localised inflammatory lesions (may ulcerate) which are precipitated by exposure to cold
18. Splinter haemorrhages	
NEUROPSYCHIATRIC	
19. Aseptic meningitis	criteria (all): acute/subacute onset headache fever abnormal CSF (raised protein &/or lymphocyte predominance) but negative cultures preferably photophobia, neck stiffness and meningeal irritation should be present as well but are not essential for diagnosis exclude CNS/meningeal infection, intracranial haemorrhage
20. Cerebral vasculitis	should be present with features of vasculitis in another system supportive imaging &/or biopsy findings
21. Demyelinating syndrome	discrete white matter lesion with associated neurological deficit not recorded elsewhere ideally there should have been at least one previously recorded event supportive imaging required exclude multiple sclerosis
22. Myelopathy	acute onset of rapidly evolving paraparesis or quadriparesis and/or sensory level exclude intramedullary and extramedullary space occupying lesion
23. Acute confusional state	acute disturbance of consciousness or level of

	arousal with reduced ability to focus, maintain or shift attention includes hypo- and hyperaroused states and encompasses the spectrum from delirium to coma
24. Psychosis	delusion or hallucinations does not occur exclusively during course of a delirium exclude drugs, substance abuse, primary psychotic disorder
25. Acute inflammatory demyelinating polyradiculoneuropathy	criteria: progressive polyradiculoneuropathy loss of reflexes symmetrical involvement increased CSF protein without pleocytosis supportive electrophysiology study
26. Mononeuropathy (single/multiplex)	supportive electrophysiology study required
27. Cranial neuropathy	except optic neuropathy which is classified under ophthalmic system
28. Plexopathy	disorder of brachial or lumbosacral plexus resulting in neurological deficit not corresponding to territory of single root or nerve supportive electrophysiology study required
29. Polyneuropathy	acute symmetrical distal sensory and/or motor deficit supportive electrophysiology study required
30. Seizure disorder	independent description of seizure by reliable witness
31. Status epilepticus	a seizure or series of seizures lasting ≥ 30 minutes without full recovery to baseline
32. Cerebrovascular disease (not due to vasculitis)	any one with supporting imaging: stroke syndrome transient ischaemic attack intracranial haemorrhage

	exclude hypoglycaemia, cerebral sinus thrombosis, vascular malformation, tumour, abscess cerebral sinus thrombosis not included as definite thrombosis not considered part of lupus activity
33. Cognitive dysfunction	significant deficits in any cognitive functions: simple attention (ability to register & maintain information) complex attention memory (ability to register, recall & recognise information eg learning, recall) visual-spatial processing (ability to analyse, synthesise & manipulate visual-spatial information) language (ability to comprehend, repeat & produce oral/written material eg verbal fluency, naming) reasoning/problem solving (ability to reason & abstract) psychomotor speed executive functions (eg planning, organising, sequencing) in absence of disturbance of consciousness or level of arousal sufficiently severe to interfere with daily activities neuropsychological testing should be done or corroborating history from third party if possible exclude substance abuse
34. Movement disorder	exclude drugs
35. Autonomic disorder	any one: fall in blood pressure to standing > 30/15 mm Hg (systolic/diastolic) increase in heart rate to standing ≥ 30 bpm loss of heart rate variation with respiration (max – min < 15 bpm, expiration:inspiration

	ratio < 1.2, Valsalva ratio < 1.4) loss of sweating over body and limbs (anhidrosis) by sweat test exclude drugs and diabetes mellitus
36. Cerebellar ataxia	cerebellar ataxia in isolation of other CNS features usually subacute presentation
37. Severe lupus headache (unremitting)	disabling headache unresponsive to narcotic analgesia & lasting ≥ 3 days exclude intracranial space occupying lesion and CNS infection
38. Headache from IC hypertension	exclude cerebral sinus thrombosis
MUSCULOSKELETAL	
39. Definite myositis	≥ 3 Bohan & Peter criteria: proximal muscle weakness elevated muscle enzymes (CK or aldolase) positive muscle biopsy EMG characteristic of myositis
40. Incomplete myositis	2 Bohan & Peter criteria
41. Severe arthritis	observed active synovitis ≥ 2 joints with marked loss of functional range of movements and significant impairment of activities of daily living, that has been present on several days (cumulatively) over the last 4 weeks
42. Arthritis or Tendonitis/Tenosynovitis	tendonitis/tenosynovitis or active synovitis ≥ 1 joint (observed or through history) with some loss of functional range of movements, that has been present on several days over the last 4 weeks
43. Arthralgia or Myalgia	inflammatory type of pain (worse in the morning with stiffness, usually improves with activity & not brought on by activity) over joints/muscle or joint tenderness which does not fulfill the above criteria for arthritis or myositis

CARDIORESPIRATORY

44. Mild myocarditis	inflammation of myocardium with raised cardiac enzymes &/or ECG changes and without resulting cardiac failure, arrhythmia or valvular dysfunction
45. Cardiac failure	cardiac failure due to myocarditis or non-infective inflammation of endocardium or cardiac valves (endocarditis) cardiac failure due to myocarditis is defined by left ventricular ejection fraction ≤ 40% & pulmonary oedema or peripheral oedema cardiac failure due to acute valvular regurgitation (from endocarditis) can be associated with normal left ventricular ejection fraction diastolic heart failure is not included
46. Arrhythmia	arrhythmia (except sinus tachycardia) due to myocarditis or non-infective inflammation of endocardium or cardiac valves (endocarditis) confirmation by electrocardiogram required (history of palpitations alone inadequate)
47. New valvular dysfunction	new cardiac valvular dysfunction due to myocarditis or non-infective inflammation of endocardium or cardiac valves (endocarditis) supportive imaging required
48. Mild serositis (pleuro-pericardial pain)	in absence of cardiac tamponade or pleural effusion with dyspnoea
49. Cardiac tamponade	supportive imaging required
50. Pleural effusion with dyspnoea	supportive imaging required
51. Pulmonary haemorrhage/vasculitis	inflammation of pulmonary vasculature with haemoptysis &/or dyspnoea &/or pulmonary hypertension supportive imaging &/or histological diagnosis required

52. Interstitial alveolitis/pneumonitis	radiological features of alveolar infiltration not due to infection or haemorrhage required for diagnosis corrected gas transfer Kco reduced to < 70% normal or fall of > 20% if previously abnormal on-going activity would be determined by clinical findings and lung function tests, and repeated imaging may be required in those with deterioration (clinically or lung function tests) or failure to respond to therapy
53. Shrinking lung syndrome	acute reduction (> 20% if previous measurement available) in lung volumes (to < 70% predicted) in the presence of normal corrected gas transfer (Kco) & dysfunctional diaphragmatic movements
54. Aortitis	inflammation of aorta (with or without dissection) with supportive imaging abnormalities accompanied by > 10 mm Hg difference in BP between arms &/or claudication of extremities &/or vascular bruits repeated imaging would be required to determine on-going activity in those with clinical deterioration or failure to respond to therapy
55. Coronary vasculitis	inflammation of coronary vessels with radiographic evidence of non-atheromatous narrowing, obstruction or aneurysmal changes

GASTROINTESTINAL

56. Lupus peritonitis	serositis presenting as acute abdomen with rebound/guarding
57. Serositis	not presenting as acute abdomen
58. Lupus enteritis or colitis	vasculitis or inflammation of small or large bowel with supportive imaging &/or biopsy findings
59. Malabsorption	diarrhoea with abnormal D- xylose absorption

	test or increased faecal fat excretion after exclusion of coeliac's disease (poor response to gluten-free diet) and gut vasculitis
60. Protein-losing enteropathy	diarrhoea with hypoalbuminaemia or increased faecal excretion of iv radiolabeled albumin after exclusion of gut vasculitis and malabsorption
61. Intestinal pseudo-obstruction	subacute intestinal obstruction due to intestinal hypomotility
62. Lupus hepatitis	raised transaminases
	absence of autoantibodies specific to autoimmune hepatitis (eg: anti-smooth muscle, anti-liver cytosol 1) &/or biopsy appearance of chronic active hepatitis
	hepatitis typically lobular with no piecemeal necrosis
63. Acute lupus cholecystitis	after exclusion of gallstones and infection
64. Acute lupus pancreatitis	usually associated multisystem involvement

OPHTHALMIC

65. Orbital inflammation	orbital inflammation with myositis &/or extra-ocular muscle swelling &/or proptosis
	supportive imaging required
66. Severe keratitis	sight threatening includes: corneal melt peripheral ulcerative keratitis
67. Mild keratitis	not sight threatening
68. Anterior uveitis	
69. Severe posterior uveitis &/or retinal vasculitis	sight-threatening &/or retinal vasculitis not due to vaso-occlusive disease
70. Mild posterior uveitis &/or retinal vasculitis	not sight-threatening
	not due to vaso-occlusive disease
71. Episcleritis	

72. Severe scleritis	necrotising anterior scleritis anterior &/or posterior scleritis requiring systemic steroids/immunosuppression &/or not responding to NSAIDs
73. Mild scleritis	anterior &/or posterior scleritis not requiring systemic steroids excludes necrotising anterior scleritis
74. Retinal/choroidal vaso-occlusive disease	includes: retinal arterial & venous occlusion serous retinal &/or retinal pigment epithelial detachments secondary to choroidal vasculopathy
75. Isolated cotton-wool spots	also known as cytoid bodies
76. Optic neuritis	excludes anterior ischaemic optic neuropathy
77. Anterior ischaemic optic neuropathy	visual loss with pale swollen optic disc due to occlusion of posterior ciliary arteries

RENAL

78. Systolic blood pressure 79. Diastolic blood pressure 80. Accelerated hypertension	blood pressure rising to > 170/110 mm Hg within 1 month with grade 3 or 4 Keith-Wagener-Barker retinal changes (flame-shaped haemorrhages or cotton-wool spots or papilloedema)
81. Urine dipstick 82. Urine albumin-creatinine ratio	on freshly voided urine sample
83. Urine protein-creatinine ratio	on freshly voided urine sample
84. 24 hour urine protein 85. Nephrotic syndrome	criteria: heavy proteinuria (> 50 mg/kg/day or > 3.5 g/day or protein-creatinine ratio > 350 mg/mmol or albumin-creatinine ratio > 350 mg/mmol) hypoalbuminaemia oedema

86. Plasma/Serum creatinine

87. GFR

MDRD formula:

GFR = 170 x [serum creatinine (mg/dl)]$^{-0.999}$ x [age]$^{-0.176}$ x [serum urea (mg/dl]$^{-0.17}$ x [serum albumin (g/dl)]$^{0.318}$ x [0.762 if female] x [1.180 if African ancestry]

units = ml/min per 1.73 m^2

normal: male = 130 ± 40
female = 120 ± 40

conversion:

serum creatinine	-	mg/dl = (μmol/l)/88.5
serum urea	-	mg/dl = (mmol/l) x 2.8
serum albumin	-	g/dl = (g/l)/10

creatinine clearance not recommended as it is not reliable

88. Active urinary sediment

pyuria (> 5 WCC/hpf or > 10 WCC/mm^3 (μl))

OR

haematuria (> 5 RBC/hpf or > 10 RBC/mm^3 (μl))

OR

red cell casts

in absence of other causes

89. Histology of active nephritis

WHO Classification (1995): (any one)
- Class III – (a) or (b) subtypes
- Class IV – (a), (b) or (c) subtypes
- Class V – (a), (b), (c) or (d) subtypes
- Vasculitis

OR

ISN/RPS Classification (2003): (any one)
- Class III – (A) or (A/C) subtypes
- Class IV – (A) or (A/C) subtypes
- Class V
- Vasculitis

within last 3 months or since previous assessments if seen less than 3 months ago

	glomerular sclerosis without inflammation not included

HAEMATOLOGICAL

90. Haemoglobin	
91. White cell count	
92. Neutrophil count	
93. Lymphocyte count	
94. Platelet count	exclude thrombocytopenia of antiphospholipid syndrome
95. TTP	thrombotic thrombocytopaenic purpura clinical syndrome of micro-angiopathic haemolytic anaemia and thrombocytopenia in absence of any other identifiable cause
96. Evidence of active haemolysis	positive Coomb's test & evidence of haemolysis (raised bilirubin or raised reticulocyte count or reduced haptoglobulins)
97. Isolated positive Coomb's test	

ADDITIONAL ITEMS

These items are required mainly for calculation of GFR

i. Weight
ii. African ancestry
iii. Serum urea
iv. Serum albumin

BILAG-2004 INDEX SCORING

- scoring based on the principle of physician's intention to treat

Grade	Definition
A	**Severe disease activity requiring any of the following treatment:** **1. systemic high dose oral corticosteroids (equivalent to prednisolone > 20 mg/day)** **2. intravenous pulse corticosteroids (equivalent to pulse methylprednisolone ≥ 500 mg)** **3. systemic immunomodulators (include biologicals, immunoglobulins and plasmapheresis)** **4. therapeutic high dose anticoagulation in the presence of high dose corticosteroids or immunomodulators** **eg: warfarin with target INR 3 - 4**
B	**Moderate disease activity requiring any of the following treatment:** **1. systemic low dose oral corticosteroids (equivalent to prednisolone ≤ 20 mg/day)** **2. intramuscular or intra-articular or soft tissue corticosteroids injection (equivalent to methylprednisolone < 500mg)** **3. topical corticosteroids** **4. topical immunomodulators** **5. antimalarials or thalidomide or prasterone or acitretin** **6. symptomatic therapy** **eg: NSAIDs for inflammatory arthritis**
C	**Stable mild disease**
D	**Inactive disease but previously affected**
E	**System never involved**

CONSTITUTIONAL

Grade A:
Pyrexia recorded as 2 (same), 3 (worse) or 4 (new) **AND**

Any 2 or more of the following recorded as 2 (same), 3 (worse) or 4 (new):

Weight loss
Lymphadenopathy/splenomegaly
Anorexia

Grade B:
Pyrexia recorded as 2 (same), 3 (worse) or 4 (new) **OR**

Any 2 or more of the following recorded as 2 (same), 3 (worse) or 4 (new):

Weight loss
Lymphadenopathy/splenomegaly
Anorexia

BUT do not fulfil criteria for Grade A

Grade C
Pyrexia recorded as 1 (improving) **OR**

One or more of the following recorded as > 0:

Weight loss
Lymphadenopathy/Splenomegaly
Anorexia

BUT does not fulfil criteria for Grade A or B

Grade D
Previous involvement

Grade E
No previous involvement

MUCOCUTANEOUS

Grade A
Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Skin eruption - severe
Angio-oedema - severe
Mucosal ulceration - severe
Panniculitis/Bullous lupus - severe
Cutaneous vasculitis/thrombosis

Grade B
Any Grade A features recorded as 1 (improving) **OR**

Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Skin eruption - mild
Panniculitis/Bullous lupus - mild
Digital infarcts/nodular vasculitis
Alopecia - severe

Grade C
Any Grade B features recorded as 1 (improving) **OR**

Any of the following recorded as > 0:

Angio-oedema - mild
Mucosal ulceration - mild
Alopecia - mild
Periungual erythema/chilblains
Splinter haemorrhages

Grade D
Previous involvement

Grade E
No previous involvement

NEUROLOGICAL

Grade A
Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Aseptic meningitis
Cerebral vasculitis
Demyelinating syndrome
Myelopathy
Acute confusional state
Psychosis
Acute inflammatory demyelinating polyradiculoneuropathy
Mononeuropathy (single/multiplex)
Cranial neuropathy
Plexopathy
Polyneuropathy
Status epilepticus
Cerebellar ataxia

Grade B
Any Grade A features recorded as 1 (improving) **OR**

Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Seizure disorder
Cerebrovascular disease (not due to vasculitis)
Cognitive dysfunction
Movement disorder
Autonomic disorder
Headache severe unremitting
Headache due to raised intracranial hypertension

Grade C
Any Grade B features recorded as 1 (improving) **OR**

Grade D
Previous involvement

Grade E
No previous involvement

MUSCULOSKELETAL

Grade A
Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Definite Myositis (Bohan and Peter)
Severe arthritis

Grade B
Any Grade A features recorded as 1 (improving) **OR**

Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Myositis with incomplete criteria
Arthritis/Tendonitis/Tenosynovitis

Grade C
Any Grade B features recorded as 1 (improving) **OR**

Any of the following recorded as > 0:

Arthralgia/Myalgia

Grade D
Previous involvement

Grade E
No previous involvement

CARDIORESPIRATORY

Grade A
Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Myocarditis/Endocarditis + Cardiac failure
Arrhythmia
New valvular dysfunction
Cardiac tamponade
Pleural effusion with dyspnoea
Pulmonary haemorrhage/vasculitis
Interstitial alveolitis/pneumonitis
Shrinking lung syndrome
Aortitis
Coronary vasculitis

Grade B
Any Grade A features recorded as 1 (improving) **OR**

Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Serositis (pleuro-pericardial pain) - mild
Myocarditis - mild

Grade C
Any Grade B features recorded as 1 (improving)

Grade D
Previous involvement

Grade E
No previous involvement

GASTROINTESTINAL

Grade A
Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Peritonitis
Lupus enteritis/colitis
Intestinal pseudo-obstruction
Acute lupus cholecystitis
Acute lupus pancreatitis

Grade B
Any Grade A feature recorded as 1 (improving) **OR**

Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Abdominal serositis and/or ascites
Malabsorption
Protein losing enteropathy
Lupus hepatitis

Grade C
Any Grade B features recorded as 1 (improving)

Grade D
Previous involvement

Grade E
No previous involvement

OPHTHALMIC

Grade A
Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Orbital inflammation/myositis/proptosis
Keratitis - severe
Posterior uveitis/retinal vasculitis - severe
Scleritis - severe
Retinal/choroidal vaso-occlusive disease
Optic neuritis
Anterior ischaemic optic neuropathy

Grade B
Any Grade A features recorded as 1 (improving) **OR**

Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Keratitis - mild
Anterior uveitis
Posterior uveitis/retinal vasculitis - mild
Scleritis - mild

Grade C
Any Grade B features recorded as 1 (improving) **OR**

Any of the following recorded as > 0:

Episcleritis
Isolated cotton-wool spots (cytoid bodies)

Grade D
Previous involvement

Grade E
No previous involvement

RENAL

Grade A
Two or more of the following **providing 1, 4 or 5 is included**:

1. Deteriorating proteinuria (severe) defined as

   (a) urine dipstick increased by ≥ 2 levels; **or**
   (b) 24 hour urine protein > 1 g that has not decreased (improved) by ≥ 50%; **or**
   (c) urine protein-creatinine ratio > 100 mg/mmol that has not decreased (improved) by ≥ 50%; **or**

   (d) urine albumin-creatinine ratio > 100 mg/mmol that has not decreased (improved) by ≥ 50%

2. Accelerated hypertension
3. Deteriorating renal function (severe) defined as

   (a) plasma creatinine > 130 μmol/l and having risen to > 130% of previous value; **or**
   (b) GFR having fallen to < 67% of previous value; **or**
   (c) GFR < 50 ml/min per 1.73 m^2, and last time was > 50 ml/min per 1.73 m^2 or was not measured.

4. Active urinary sediment
5. Histological evidence of active nephritis within last 3 months
6. Nephrotic syndrome

Grade B
One of the following:

1. One of the Grade A feature

2. Urine dipstick which has risen by 1 level to at least 2+

3. Plasma creatinine > 130 μmol/l and having risen to ≥ 115% but ≤ 130% of previous value

Grade C
One of the following:

1. Mild/Stable proteinuria defined as

   (a) urine dipstick ≥ 1+ but has not fulfilled criteria for Grade A & B; **or**
   (b) 24 hour urine protein > 0.25 g but has not fulfilled criteria for Grade A & B ; **or**
   (c) urine protein-creatinine ratio > 25 mg/mmol but has not fulfilled criteria for Grade A & B; **or**

   (d) urine albumin-creatinine ratio > 25 mg/mmol but has not fulfilled criteria for Grade A

& B

2. Rising blood pressure (providing the recorded values are > 140/90 mm Hg) which has not fulfilled criteria for Grade A & B, defined as

(a) systolic rise of ≥ 30 mm Hg; **and**
(b) diastolic rise of ≥ 15mm Hg

Grade D
Previous involvement

Grade E
No previous involvement

Note: although albumin-creatinine ratio and protein-creatinine ratio are different, the same cut-off values are used for both in this index

HAEMATOLOGICAL

Grade A
TTP recorded as 2 (same), 3 (worse) or 4 (new) **OR**

Any of the following:

Haemoglobin	< 8 g/dl
White cell count	< 1.0 x 10^9/l
Neutrophil count	< 0.5 x 10^9/l
Platelet count	< 25 x 10^9/l

Grade B
TTP recorded as 1 (improving) **OR**

Any of the following:

Haemoglobin	8 - 9.9 x 10^9/l
White cell count	1 - 2.4 x 10^9/l
Neutrophil count	0.5 - 1.4 x 10^9/l
Platelet count	25 - 49 x 10^9/l
Evidence of active haemolysis	

Grade C
Any of the following:

Haemoglobin	10 - 10.9 g/dl
White cell count	2.5 - 3.9 x 10^9/l
Neutrophil count	1.5 - 1.9 x 10^9/l
Lymphocyte count	< 1.0 x 10^9/L
Platelet count	50 - 149 x 10^9/l
Isolated Coombs' test positive	

Grade D
Previous involvement

Grade E
No previous involvement

Appendix 7 Revised BILAG-2004 Index (Post-Longitudinal Study)

BILAG-2004 INDEX

Only record items due to SLE Disease Activity & assessment refers to manifestations occurring in the last 4 weeks (compared with the previous 4 weeks).

♦♦ TO BE USED WITH THE GLOSSARY ♦♦

Scoring:
ND Not Done
1 Improving
2 Same
3 Worse
4 New
Yes/No OR Value (where indicated)
❑ indicate if not due to SLE activity
(default is 0 = not present)

CONSTITUTIONAL

1. Pyrexia - documented > 37.5ºC	(	)
2. Weight loss - unintentional > 5%	(	)
3. Lymphadenopathy/splenomegaly	(	)
4. Anorexia	(	)

MUCOCUTANEOUS

5. Skin eruption - severe	(	)
6. Skin eruption - mild	(	)
7. Angio-oedema - severe	(	)
8. Angio-oedema - mild	(	)
9. Mucosal ulceration - severe	(	)
10. Mucosal ulceration - mild	(	)
11. Panniculitis/Bullous lupus - severe	(	)
12. Panniculitis/Bullous lupus - mild	(	)
13. Cutaneous vasculitis/thrombosis	(	)
14. Digital infarcts/nodular vasculitis	(	)
15. Alopecia - severe	(	)
16. Alopecia - mild	(	)
17. Peri-ungual erythema/chilblains	(	)
18. Splinter haemorrhages	(	)

NEUROPSYCHIATRIC

19. Aseptic meningitis	(	)
20. Cerebral vasculitis	(	)
21. Demyelinating syndrome	(	)
22. Myelopathy	(	)
23. Acute confusional state	(	)
24. Psychosis	(	)
25. Acute inflammatory demyelinating polyradiculoneuropathy	(	)
26. Mononeuropathy (single/multiplex)	(	)

27. Cranial neuropathy ( )
28. Plexopathy ( )
29. Polyneuropathy ( )
30. Seizure disorder ( )
31. Status epilepticus ( )
32. Cerebrovascular disease (not due to vasculitis) ( )
33. Cognitive dysfunction ( )
34. Movement disorder ( )
35. Autonomic disorder ( )
36. Cerebellar ataxia (isolated) ( )
37. Lupus headache - severe unremitting ( )
38. Headache from IC hypertension ( )

MUSCULOSKELETAL
39. Myositis - severe ( )
40. Myositis - mild ( )
41. Arthritis(severe) ( )
42. Arthritis(moderate)/Tendonitis/Tenosynovitis ( )
43. Arthritis(mild)/Arthralgia/Myalgia ( )

CARDIORESPIRATORY
44. Myocarditis - mild ( )
45. Myocarditis/Endocarditis + Cardiac failure ( )
46. Arrhythmia ( )
47. New valvular dysfunction ( )
48. Pleurisy/Pericarditis ( )
49. Cardiac tamponade ( )
50. Pleural effusion with dyspnoea ( )
51. Pulmonary haemorrhage/vasculitis ( )
52. Interstitial alveolitis/pneumonitis ( )
53. Shrinking lung syndrome ( )
54. Aortitis ( )
55. Coronary vasculitis ( )

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 ( )

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		( )	

RENAL

78. Systolic blood pressure (mm Hg)	value	( )	❑
79. Diastolic blood pressure (mm Hg)	value	( )	❑
80. Accelerated hypertension	Yes/No	( )	
81. Urine dipstick protein	(+=1, ++=2, +++=3)	( )	❑
82. Urine albumin-creatinine ratio	mg/mmol	( )	❑
83. Urine protein-creatinine ratio	mg/mmol	( )	❑
84. 24 hour urine protein (g)	value	( )	❑
85. Nephrotic syndrome	Yes/No	( )	
86. Creatinine (plasma/serum)	µmol/l	( )	❑
87. GFR (calculated)	ml/min/1.73 m^2	( )	❑
88. Active urinary sediment	Yes/No	( )	
89. Active nephritis	Yes/No	( )	

HAEMATOLOGICAL

90. Haemoglobin (g/dl)	value	( )	❑
91. Total white cell count (x 10^9/l)	value	( )	❑
92. Neutrophils (x 10^9/l)	value	( )	❑
93. Lymphocytes (x 10^9/l)	value	( )	❑
94. Platelets (x 10^9/l)	value	( )	❑
95. TTP		( )	
96. Evidence of active haemolysis	Yes/No	( )	
97. Coombs' test positive (isolated)	Yes/No	( )	

BILAG-2004 INDEX GLOSSARY

INSTRUCTIONS

- only record features that are **attributable to SLE disease activity and not due to damage, infection, thrombosis (in absence of inflammatory process) or other conditions**

- assessment refers to manifestations occurring in the **last 4 weeks compared with the previous 4 weeks**

- activity refers to disease process which is reversible while damage refers to permanent process/scarring (irreversible)

- damage due to SLE should be considered as a cause of features that are fixed/persistent (SLICC/ACR damage index uses persistence ≥ 6 months to define damage)

- in some manifestations, it may be difficult to differentiate SLE from other conditions as there may not be any specific test and the decision would then lies with the **physician's judgement on the balance of probabilities**

- ophthalmic manifestations would usually need to be assessed by ophthalmologist and these items would need to be recorded retrospectively after receiving the response from ophthalmologist

- guidance for scoring:

  (4) NEW
  - manifestations are recorded as new when it is a new episode occurring in the last 4 weeks (compared to the previous 4 weeks) which has not improved and this includes new episodes (recurrence) of old manifestations

  - new episode occurring in the last 4 weeks but also satisfying the criteria for improvement (below) would be classified as improving instead of new

  (3) WORSE
  - this refers to manifestations that have deteriorated in the last 4 weeks compared to the previous 4 weeks

  (2) SAME
  - this refers to manifestations that have been present for the last 4 weeks and the previous 4 weeks without significant improvement or deterioration (from the previous 4 weeks)

  - this also applies to manifestations that have improved over the last 4 weeks compared to the previous 4 weeks but do not meet the criteria for improvement

(1) IMPROVING

- definition of **improvement**: (a) the amount of improvement is sufficient for **consideration of reduction in therapy** and would not justify escalation in therapy

AND

(b) improvement must be **present currently and for at least 2 weeks** out of the last 4 weeks

OR

manifestation that has **completely resolved and remained absent** over the **whole of last 1 week**

(0) NOT PRESENT

(ND) NOT DONE

- it is important to indicate if a test has not been performed (particularly laboratory investigations) so that this will be recorded as such in the database & not as normal or absent (which is the default)

❑ INDICATE (TICK) IF NOT DUE TO SLE ACTIVITY

- for descriptors that are based on measurements (in renal and haematology systems), it is important to indicate if these are not due to lupus disease activity (for consideration of scoring) as they are usually recorded routinely into a database

CHANGE IN SEVERITY CATEGORY

- there are several items in the index which have been divided into categories of mild and severe (depending on definition). It is essential to record mild and severe items appropriately if the manifestations fulfill both criteria during the last 4 weeks

- if a mild item deteriorated to the extent that it fulfilled the definition of severe category (ie changed into severe category) within the last 4 weeks:
  severe item scored as new (4)
  AND mild item scored as worsening (3)

- if a severe item improved (fulfilling the improvement criteria) to the extent that it no longer fulfilled the definition of severe category (ie changed into mild category) within the last 4 weeks:
  severe item scored as not present (0) if criteria for severe category has not been met over last 4 weeks
  or as improving (1) if criteria for severe category has been met at some point over last 4 weeks

AND

mild item scored as improving (1) if it is improving over last 4 weeks
or as the same (2) if it has remained stable over last 4 weeks

CONSTITUTIONAL

1. Pyrexia — temperature > 37.5˚C documented

2. Unintentional weight loss > 5%

3. Lymphadenopathy — lymph node more than 1 cm diameter

exclude infection

4. Anorexia

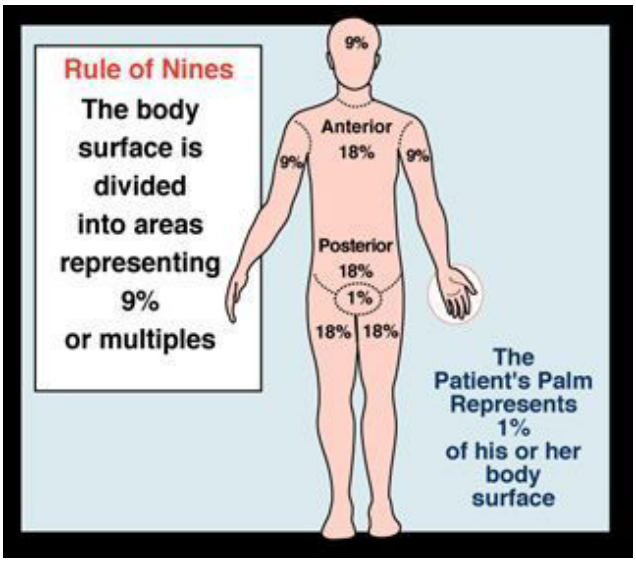
MUCOCUTANEOUS

5. Severe eruption — > 18% body surface area

any lupus rash except panniculitis, bullous lesion & angio-oedema

body surface area (BSA) is estimated using the rules of nines (used to assess extent of burns) as follows:

palm(excluding fingers) = 1% BSA
each lower limb = 18% BSA
each upper limb = 9% BSA
torso (front) = 18% BSA
torso (back) = 18% BSA
head = 9% BSA
genital (male) = 1% BSA



6. Mild eruption — ≤ 18% body surface area

	any lupus rash except panniculitis, bullous lesion & angio-oedema malar rash must have been observed by a physician and has to be present continuously (persistent) for at least 1 week to be considered significant (to be recorded)
7. Severe angio-oedema	potentially life-threatening eg: stridor angio-oedema is a variant form of urticaria which affects the subcutaneous, submucosal and deep dermal tissues
8. Mild angio-oedema	not life threatening
9. Severe mucosal ulceration	disabling (significantly interfering with oral intake), extensive & deep ulceration must have been observed by a physician
10. Mild mucosal ulceration	localised &/or non-disabling ulceration
11. Severe panniculitis or bullous lupus	any one: > 9% body surface area facial panniculitis panniculitis that is beginning to ulcerate panniculitis that threatens integrity of subcutaneous tissue (beginning to cause surface depression) on > 9% body surface area panniculitis presents as a palpable and tender subcutaneous induration/nodule note that established surface depression and atrophy alone is likely to be due to damage
12. Mild panniculitis or bullous lupus	≤ 9% body surface area does not fulfill any criteria for severe panniculitis (for panniculitis)
13. Major cutaneous vasculitis/thrombosis	resulting in extensive gangrene or ulceration or skin infarction
14. Digital infarct or nodular vasculitis	localised single or multiple infarct(s) over digit(s) or tender erythematous nodule(s)

15. Severe alopecia	clinically detectable (diffuse or patchy) hair loss with scalp inflammation (redness over scalp)
16. Mild alopecia	diffuse or patchy hair loss without scalp inflammation (clinically detectable or by history)
17. Peri-ungual erythema or chilblains	chilblains are localised inflammatory lesions (may ulcerate) which are precipitated by exposure to cold
18. Splinter haemorrhages	

NEUROPSYCHIATRIC

19. Aseptic meningitis	criteria (all): acute/subacute onset headache fever abnormal CSF (raised protein &/or lymphocyte predominance) but negative cultures preferably photophobia, neck stiffness and meningeal irritation should be present as well but are not essential for diagnosis exclude CNS/meningeal infection, intracranial haemorrhage
20. Cerebral vasculitis	should be present with features of vasculitis in another system supportive imaging &/or biopsy findings
21. Demyelinating syndrome	discrete white matter lesion with associated neurological deficit not recorded elsewhere ideally there should have been at least one previously recorded event supportive imaging required exclude multiple sclerosis
22. Myelopathy	acute onset of rapidly evolving paraparesis or quadriparesis and/or sensory level exclude intramedullary and extramedullary

	space occupying lesion
23. Acute confusional state	acute disturbance of consciousness or level of arousal with reduced ability to focus, maintain or shift attention includes hypo- and hyperaroused states and encompasses the spectrum from delirium to coma
24. Psychosis	delusion or hallucinations does not occur exclusively during course of a delirium exclude drugs, substance abuse, primary psychotic disorder
25. Acute inflammatory demyelinating polyradiculoneuropathy	criteria: progressive polyradiculoneuropathy loss of reflexes symmetrical involvement increased CSF protein without pleocytosis supportive electrophysiology study
26. Mononeuropathy (single/multiplex)	supportive electrophysiology study required
27. Cranial neuropathy	except optic neuropathy which is classified under ophthalmic system
28. Plexopathy	disorder of brachial or lumbosacral plexus resulting in neurological deficit not corresponding to territory of single root or nerve supportive electrophysiology study required
29. Polyneuropathy	acute symmetrical distal sensory and/or motor deficit supportive electrophysiology study required
30. Seizure disorder	independent description of seizure by reliable witness
31. Status epilepticus	a seizure or series of seizures lasting ≥ 30 minutes without full recovery to baseline
32. Cerebrovascular disease	any one with supporting imaging:

(not due to vasculitis)	stroke syndrome transient ischaemic attack intracranial haemorrhage exclude hypoglycaemia, cerebral sinus thrombosis, vascular malformation, tumour, abscess cerebral sinus thrombosis not included as definite thrombosis not considered part of lupus activity
33. Cognitive dysfunction	significant deficits in any cognitive functions: simple attention (ability to register & maintain information) complex attention memory (ability to register, recall & recognise information eg learning, recall) visual-spatial processing (ability to analyse, synthesise & manipulate visual-spatial information) language (ability to comprehend, repeat & produce oral/written material eg verbal fluency, naming) reasoning/problem solving (ability to reason & abstract) psychomotor speed executive functions (eg planning, organising, sequencing) in absence of disturbance of consciousness or level of arousal sufficiently severe to interfere with daily activities neuropsychological testing should be done or corroborating history from third party if possible exclude substance abuse
34. Movement disorder	exclude drugs
35. Autonomic disorder	any one: fall in blood pressure to standing > 30/15 mm Hg (systolic/diastolic) increase in heart rate to standing ≥ 30 bpm

	loss of heart rate variation with respiration (max – min < 15 bpm, expiration:inspiration ratio < 1.2, Valsalva ratio < 1.4)
	loss of sweating over body and limbs (anhidrosis) by sweat test
	exclude drugs and diabetes mellitus
36. Cerebellar ataxia	cerebellar ataxia in isolation of other CNS features
	usually subacute presentation
37. Severe lupus headache (unremitting)	disabling headache unresponsive to narcotic analgesia & lasting ≥ 3 days
	exclude intracranial space occupying lesion and CNS infection
38. Headache from IC hypertension	exclude cerebral sinus thrombosis
MUSCULOSKELETAL	
39. Severe myositis	significantly elevated serum muscle enzymes with significant muscle weakness
	exclude endocrine causes and drug-induced myopathy
	electromyography and muscle biopsy are used for diagnostic purpose and are not required to determine level of activity
40. Mild myositis	significantly elevated serum muscle enzymes with myalgia but without significant muscle weakness
	asymptomatic elevated serum muscle enzymes not included
	exclude endocrine causes and drug-induced myopathy

	electromyography and muscle biopsy are used for diagnostic purpose and are not required to determine level of activity
41. Severe arthritis	observed active synovitis ≥ 2 joints with marked loss of functional range of movements and significant impairment of activities of daily living, that has been present on several days (cumulatively) over the last 4 weeks
42. Moderate arthritis or Tendonitis or Tenosynovitis	tendonitis/tenosynovitis or active synovitis ≥ 1 joint (observed or through history) with some loss of functional range of movements, that has been present on several days over the last 4 weeks
43. Mild arthritis or Arthralgia or Myalgia	inflammatory type of pain (worse in the morning with stiffness, usually improves with activity & not brought on by activity) over joints/muscle inflammatory arthritis which does not fulfill the above criteria for moderate or severe arthritis
CARDIORESPIRATORY	
44. Mild myocarditis	inflammation of myocardium with raised cardiac enzymes &/or ECG changes and without resulting cardiac failure, arrhythmia or valvular dysfunction
45. Cardiac failure	cardiac failure due to myocarditis or non-infective inflammation of endocardium or cardiac valves (endocarditis) cardiac failure due to myocarditis is defined by left ventricular ejection fraction ≤ 40% & pulmonary oedema or peripheral oedema cardiac failure due to acute valvular regurgitation (from endocarditis) can be associated with normal left ventricular ejection fraction diastolic heart failure is not included
46. Arrhythmia	arrhythmia (except sinus tachycardia) due to myocarditis or non-infective inflammation of endocardium or cardiac valves (endocarditis)

	confirmation by electrocardiogram required (history of palpitations alone inadequate)
47. New valvular dysfunction	new cardiac valvular dysfunction due to myocarditis or non-infective inflammation of endocardium or cardiac valves (endocarditis) supportive imaging required
48. Pleurisy/Pericarditis	convincing history &/or physical findings that you would consider treating in absence of cardiac tamponade or pleural effusion with dyspnoea do not score if you are unsure whether or not it is pleurisy/pericarditis
49. Cardiac tamponade	supportive imaging required
50. Pleural effusion with dyspnoea	supportive imaging required
51. Pulmonary haemorrhage/vasculitis	inflammation of pulmonary vasculature with haemoptysis &/or dyspnoea &/or pulmonary hypertension supportive imaging &/or histological diagnosis required
52. Interstitial alveolitis/pneumonitis	radiological features of alveolar infiltration not due to infection or haemorrhage required for diagnosis corrected gas transfer Kco reduced to < 70% normal or fall of > 20% if previously abnormal on-going activity would be determined by clinical findings and lung function tests, and repeated imaging may be required in those with deterioration (clinically or lung function tests) or failure to respond to therapy
53. Shrinking lung syndrome	acute reduction (> 20% if previous measurement available) in lung volumes (to < 70% predicted) in the presence of normal corrected gas transfer (Kco) & dysfunctional diaphragmatic movements
54. Aortitis	inflammation of aorta (with or without

	dissection) with supportive imaging abnormalities
	accompanied by > 10 mm Hg difference in BP between arms &/or claudication of extremities &/or vascular bruits
	repeated imaging would be required to determine on-going activity in those with clinical deterioration or failure to respond to therapy
55. Coronary vasculitis	inflammation of coronary vessels with radiographic evidence of non-atheromatous narrowing, obstruction or aneurysmal changes

GASTROINTESTINAL

56. Lupus peritonitis	serositis presenting as acute abdomen with rebound/guarding
57. Serositis	not presenting as acute abdomen
58. Lupus enteritis or colitis	vasculitis or inflammation of small or large bowel with supportive imaging &/or biopsy findings
59. Malabsorption	diarrhoea with abnormal D- xylose absorption test or increased faecal fat excretion after exclusion of coeliac's disease (poor response to gluten-free diet) and gut vasculitis
60. Protein-losing enteropathy	diarrhoea with hypoalbuminaemia or increased faecal excretion of iv radiolabeled albumin after exclusion of gut vasculitis and malabsorption
61. Intestinal pseudo-obstruction	subacute intestinal obstruction due to intestinal hypomotility
62. Lupus hepatitis	raised transaminases
	absence of autoantibodies specific to autoimmune hepatitis (eg: anti-smooth muscle, anti-liver cytosol 1) &/or biopsy appearance of chronic active hepatitis
	hepatitis typically lobular with no piecemeal necrosis

	exclude drug-induced and viral hepatitis
63. Acute lupus cholecystitis	after exclusion of gallstones and infection
64. Acute lupus pancreatitis	usually associated multisystem involvement

OPHTHALMIC

65. Orbital inflammation	orbital inflammation with myositis &/or extra-ocular muscle swelling &/or proptosis supportive imaging required
66. Severe keratitis	sight threatening includes: corneal melt peripheral ulcerative keratitis
67. Mild keratitis	not sight threatening
68. Anterior uveitis	
69. Severe posterior uveitis &/or retinal vasculitis	sight-threatening &/or retinal vasculitis not due to vaso-occlusive disease
70. Mild posterior uveitis &/or retinal vasculitis	not sight-threatening not due to vaso-occlusive disease
71. Episcleritis	
72. Severe scleritis	necrotising anterior scleritis anterior &/or posterior scleritis requiring systemic steroids/immunosuppression &/or not responding to NSAIDs
73. Mild scleritis	anterior &/or posterior scleritis not requiring systemic steroids excludes necrotising anterior scleritis
74. Retinal/choroidal vaso-occlusive disease	includes: retinal arterial & venous occlusion serous retinal &/or retinal pigment epithelial detachments secondary to choroidal vasculopathy
75. Isolated cotton-wool spots	also known as cytoid bodies

76. Optic neuritis	excludes anterior ischaemic optic neuropathy
77. Anterior ischaemic optic neuropathy	visual loss with pale swollen optic disc due to occlusion of posterior ciliary arteries

RENAL

78. Systolic blood pressure	
79. Diastolic blood pressure	
80. Accelerated hypertension	blood pressure rising to > 170/110 mm Hg within 1 month with grade 3 or 4 Keith-Wagener-Barker retinal changes (flame-shaped haemorrhages or cotton-wool spots or papilloedema)
81. Urine dipstick	
82. Urine albumin-creatinine ratio	on freshly voided urine sample conversion: 1 mg/mg = 113 mg/mmol it is important to exclude other causes (especially infection) when proteinuria is present
83. Urine protein-creatinine ratio	on freshly voided urine sample conversion: 1 mg/mg = 113 mg/mmol it is important to exclude other causes (especially infection) when proteinuria is present
84. 24 hour urine protein	it is important to exclude other causes (especially infection) when proteinuria is present
85. Nephrotic syndrome	criteria: heavy proteinuria (≥ 3.5 g/day or protein-creatinine ratio ≥ 350 mg/mmol or albumin-creatinine ratio ≥ 350 mg/mmol) hypoalbuminaemia oedema
86. Plasma/Serum creatinine	
87. GFR	MDRD formula: GFR = 170 x [serum creatinine (mg/dl)]$^{-0.999}$ x [age]$^{-0.176}$ x [serum urea (mg/dl]$^{-0.17}$ x [serum albumin (g/dl)]$^{0.318}$ x [0.762 if female] x [1.180 if African ancestry]

units = ml/min per 1.73 m^2
normal: male = 130 ± 40
female = 120 ± 40

conversion:
serum creatinine - mg/dl = (μmol/l)/88.5
serum urea - mg/dl = (mmol/l) x 2.8
serum albumin - g/dl = (g/l)/10

creatinine clearance not recommended as it is not reliable

exclude other causes for decrease in GFR (especially drugs)

88. Active urinary sediment

pyuria (> 5 WCC/hpf or > 10 WCC/mm^3 (μl))

OR

haematuria (> 5 RBC/hpf or > 10 RBC/mm^3 (μl))

OR

red cell casts

OR

white cell casts

in absence of other causes (especially infection, vaginal bleed, calculi)

89. Histology of active nephritis

WHO Classification (1995): (any one)
Class III – (a) or (b) subtypes
Class IV – (a), (b) or (c) subtypes
Class V – (a), (b), (c) or (d) subtypes
Vasculitis

OR

ISN/RPS Classification (2003): (any one)
Class III – (A) or (A/C) subtypes
Class IV – (A) or (A/C) subtypes
Class V
Vasculitis

within last 3 months

	glomerular sclerosis without inflammation not included

HAEMATOLOGICAL

90. Haemoglobin	exclude dietary deficiency & GI blood loss
91. White cell count	exclude drug-induced cause
92. Neutrophil count	exclude drug-induced cause
93. Lymphocyte count	
94. Platelet count	exclude thrombocytopenia of antiphospholipid syndrome & drug-induced cause
95. TTP	thrombotic thrombocytopaenic purpura clinical syndrome of micro-angiopathic haemolytic anaemia and thrombocytopenia in absence of any other identifiable cause
96. Evidence of active haemolysis	positive Coomb's test & evidence of haemolysis (raised bilirubin or raised reticulocyte count or reduced haptoglobulins)
97. Isolated positive Coomb's test	

ADDITIONAL ITEMS

These items are required mainly for calculation of GFR

i. Weight
ii. African ancestry
iii. Serum urea
iv. Serum albumin

BILAG-2004 INDEX SCORING

- scoring based on the principle of physician's intention to treat

Grade	Definition
A	**Severe disease activity requiring any of the following treatment:** **1. systemic high dose oral corticosteroids (equivalent to prednisolone > 20 mg/day)** **2. intravenous pulse corticosteroids (equivalent to pulse methylprednisolone ≥ 500 mg)** **3. systemic immunomodulators (include biologicals, immunoglobulins and plasmapheresis)** **4. therapeutic high dose anticoagulation in the presence of high dose corticosteroids or immunomodulators** **eg: warfarin with target INR 3 - 4**
B	**Moderate disease activity requiring any of the following treatment:** **1. systemic low dose oral corticosteroids (equivalent to prednisolone ≤ 20 mg/day)** **2. intramuscular or intra-articular or soft tissue corticosteroids injection (equivalent to methylprednisolone < 500mg)** **3. topical corticosteroids** **4. topical immunomodulators** **5. antimalarials or thalidomide or prasterone or acitretin** **6. symptomatic therapy** **eg: NSAIDs for inflammatory arthritis**
C	**Mild disease**
D	**Inactive disease but previously affected**
E	**System never involved**

CONSTITUTIONAL

Grade A:
Pyrexia recorded as 2 (same), 3 (worse) or 4 (new) **AND**

Any 2 or more of the following recorded as 2 (same), 3 (worse) or 4 (new):

Weight loss
Lymphadenopathy/splenomegaly
Anorexia

Grade B:
Pyrexia recorded as 2 (same), 3 (worse) or 4 (new) **OR**

Any 2 or more of the following recorded as 2 (same), 3 (worse) or 4 (new):

Weight loss
Lymphadenopathy/splenomegaly
Anorexia

BUT do not fulfil criteria for Grade A

Grade C
Pyrexia recorded as 1 (improving) **OR**

One or more of the following recorded as > 0:

Weight loss
Lymphadenopathy/Splenomegaly
Anorexia

BUT does not fulfil criteria for Grade A or B

Grade D
Previous involvement

Grade E
No previous involvement

MUCOCUTANEOUS

Grade A
Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Skin eruption - severe
Angio-oedema - severe
Mucosal ulceration - severe
Panniculitis/Bullous lupus - severe
Major cutaneous vasculitis/thrombosis

Grade B
Any Grade A features recorded as 1 (improving) **OR**

Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Skin eruption - mild
Panniculitis/Bullous lupus - mild
Digital infarcts or nodular vasculitis
Alopecia - severe

Grade C
Any Grade B features recorded as 1 (improving) **OR**

Any of the following recorded as > 0:

Angio-oedema - mild
Mucosal ulceration - mild
Alopecia - mild
Periungual erythema/chilblains
Splinter haemorrhages

Grade D
Previous involvement

Grade E
No previous involvement

NEUROPSYCHIATRIC

Grade A
Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Aseptic meningitis
Cerebral vasculitis
Demyelinating syndrome
Myelopathy
Acute confusional state
Psychosis
Acute inflammatory demyelinating polyradiculoneuropathy
Mononeuropathy (single/multiplex)
Cranial neuropathy
Plexopathy
Polyneuropathy
Status epilepticus
Cerebellar ataxia

Grade B
Any Grade A features recorded as 1 (improving) **OR**

Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Seizure disorder
Cerebrovascular disease (not due to vasculitis)
Cognitive dysfunction
Movement disorder
Autonomic disorder
Lupus headache - severe unremitting
Headache due to raised intracranial hypertension

Grade C
Any Grade B features recorded as 1 (improving)

Grade D
Previous involvement

Grade E
No previous involvement

MUSCULOSKELETAL

Grade A
Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Severe Myositis
Severe Arthritis

Grade B
Any Grade A features recorded as 1 (improving) **OR**

Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Mild Myositis
Moderate Arthritis/Tendonitis/Tenosynovitis

Grade C
Any Grade B features recorded as 1 (improving) **OR**

Any of the following recorded as > 0:

Mild Arthritis/Arthralgia/Myalgia

Grade D
Previous involvement

Grade E
No previous involvement

CARDIORESPIRATORY

Grade A
Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Myocarditis/Endocarditis + Cardiac failure
Arrhythmia
New valvular dysfunction
Cardiac tamponade
Pleural effusion with dyspnoea
Pulmonary haemorrhage/vasculitis
Interstitial alveolitis/pneumonitis
Shrinking lung syndrome
Aortitis
Coronary vasculitis

Grade B
Any Grade A features recorded as 1 (improving) **OR**

Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Pleurisy/Pericarditis
Myocarditis - mild

Grade C
Any Grade B features recorded as 1 (improving)

Grade D
Previous involvement

Grade E
No previous involvement

GASTROINTESTINAL

Grade A
Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Peritonitis
Lupus enteritis/colitis
Intestinal pseudo-obstruction
Acute lupus cholecystitis
Acute lupus pancreatitis

Grade B
Any Grade A feature recorded as 1 (improving) **OR**

Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Abdominal serositis and/or ascites
Malabsorption
Protein losing enteropathy
Lupus hepatitis

Grade C
Any Grade B features recorded as 1 (improving)

Grade D
Previous involvement

Grade E
No previous involvement

OPHTHALMIC

Grade A
Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Orbital inflammation/myositis/proptosis
Keratitis - severe
Posterior uveitis/retinal vasculitis - severe
Scleritis - severe
Retinal/choroidal vaso-occlusive disease
Optic neuritis
Anterior ischaemic optic neuropathy

Grade B
Any Grade A features recorded as 1 (improving) **OR**

Any of the following recorded as 2 (same), 3 (worse) or 4 (new):

Keratitis - mild
Anterior uveitis
Posterior uveitis/retinal vasculitis - mild
Scleritis - mild

Grade C
Any Grade B features recorded as 1 (improving) **OR**

Any of the following recorded as > 0:

Episcleritis
Isolated cotton-wool spots (cytoid bodies)

Grade D
Previous involvement

Grade E
No previous involvement

RENAL

Grade A
Two or more of the following **providing 1, 4 or 5 is included**:

1. Deteriorating proteinuria (severe) defined as

   (a) urine dipstick increased by ≥ 2 levels (used only if other methods of urine protein estimation not available); **or**

   (b) 24 hour urine protein > 1 g that has not decreased (improved) by ≥ 25%; **or**
   (c) urine protein-creatinine ratio > 100 mg/mmol that has not decreased (improved) by ≥ 25%; **or**

   (d) urine albumin-creatinine ratio > 100 mg/mmol that has not decreased (improved) by ≥ 25%

2. Accelerated hypertension
3. Deteriorating renal function (severe) defined as

   (a) plasma creatinine > 130 μmol/l and having risen to > 130% of previous value; **or**
   (b) GFR < 80 ml/min per 1.73 m^2 and having fallen to < 67% of previous value; **or**
   (c) GFR < 50 ml/min per 1.73 m^2, and last time was > 50 ml/min per 1.73 m^2 or was not measured.

4. Active urinary sediment
5. Histological evidence of active nephritis within last 3 months
6. Nephrotic syndrome

Grade B
One of the following:

1. One of the Grade A feature

2. Proteinuria (that has not fulfilled Grade A criteria)
   (a) urine dipstick which has risen by 1 level to at least 2+ (used only if other methods of urine protein estimation not available); **or**

   (b) 24 hour urine protein ≥ 0.5 g that has not decreased (improved) by ≥ 25%; **or**
   (c) urine protein-creatinine ratio ≥ 50 mg/mmol that has not decreased (improved) by ≥ 25%; **or**

   (d) urine albumin-creatinine ratio ≥ 50 mg/mmol that has not decreased (improved) by ≥ 25%

3. Plasma creatinine > 130 μmol/l and having risen to ≥ 115% but ≤ 130% of previous value

Grade C
One of the following:

1. Mild/Stable proteinuria defined as

(a) urine dipstick ≥ 1+ but has not fulfilled criteria for Grade A & B (used only if other methods of urine protein estimation not available); **or**

(b) 24 hour urine protein > 0.25 g but has not fulfilled criteria for Grade A & B ; **or**
(c) urine protein-creatinine ratio > 25 mg/mmol but has not fulfilled criteria for Grade A & B; **or**

(d) urine albumin-creatinine ratio > 25 mg/mmol but has not fulfilled criteria for Grade A & B

2. Rising blood pressure (providing the recorded values are > 140/90 mm Hg) which has not fulfilled criteria for Grade A & B, defined as

(a) systolic rise of ≥ 30 mm Hg; **and**
(b) diastolic rise of ≥ 15mm Hg

Grade D
Previous involvement

Grade E
No previous involvement

Note: although albumin-creatinine ratio and protein-creatinine ratio are different, the same cut-off values are used for both in this index

HAEMATOLOGICAL

Grade A
TTP recorded as 2 (same), 3 (worse) or 4 (new) **OR**

Any of the following:

Haemoglobin	< 8 g/dl
White cell count	< 1.0 x 10^9/l
Neutrophil count	< 0.5 x 10^9/l
Platelet count	< 25 x 10^9/l

Grade B
TTP recorded as 1 (improving) **OR**

Any of the following:

Haemoglobin	8 - 8.9 g/dl
White cell count	1 - 1.9 x 10^9/l
Neutrophil count	0.5 - 0.9 x 10^9/l
Platelet count	25 - 49 x 10^9/l
Evidence of active haemolysis	

Grade C
Any of the following:

Haemoglobin	9 - 10.9 g/dl
White cell count	2 - 3.9 x 10^9/l
Neutrophil count	1 - 1.9 x 10^9/l
Lymphocyte count	< 1.0 x 10^9/L
Platelet count	50 - 149 x 10^9/l
Isolated Coombs' test positive	

Grade D
Previous involvement

Grade E
No previous involvement

Publications

Papers arising directly from the work presented in this thesis and published at the time of submission of the thesis:

Yee CS, Farewell V, Isenberg DA, Prabu A, Sokoll K, Teh LS, Rahman A, Bruce IN, Griffiths B, Akil M, McHugh N, D'Cruz D, Khamashta MA, Bowman S, Maddison P, Zoma A, Allen E, Gordon C. Revised BILAG-2004 index: A reliable tool for assessment of systemic lupus erythematosus activity. Arthritis & Rheumatism 2006;54(10):3300-5.

Yee CS, Farewell V, Isenberg DA, Rahman A, Teh LS, Griffiths B, Bruce IN, Ahmad Y, Prabu A, Akil M, McHugh N, D'Cruz D, Khamashta MA, Maddison P, Gordon C.BILAG-2004 Index is valid for assessment of SLE disease activity. Arthritis and Rheumatism 2007;56(12):4113-9.

Yee CS, Isenberg DA, Prabu A, Sokoll K, Teh LS, Rahman A, Bruce IN, Griffiths B, Akil M, McHugh N, D'Cruz D, Khamashta MA, Maddison P, Zoma A, Gordon C. BILAG-2004 Index captures SLE disease activity better than SLEDAI-2000. Annals of Rheumatic Diseases 2008;67(6):873-6.

References

References

Abrahamowicz M., Fortin P. R., Du B. R., Nayak V., Neville C., & Liang M. H. 1998, "The relationship between disease activity and expert physician's decision to start major treatment in active systemic lupus erythematosus: a decision aid for development of entry criteria for clinical trials", *J.Rheumatol.*, vol. 25, no. 2, pp. 277-284.

Abu-Shakra M., Urowitz M. B., Gladman D. D., & Gough J. 1995, "Mortality studies in systemic lupus erythematosus. Results from a single center. I. Causes of death", *J.Rheumatol.*, vol. 22, no. 7, pp. 1259-1264.

ACR Ad Hoc Committee on neuropsychiatric lupus nomenclature 1999, "The American College of Rheumatology nomenclature and case definitions for neuropsychiatric lupus syndromes", *Arthritis Rheum.*, vol. 42, no. 4, pp. 599-608.

Alarcon G. S., Friedman A. W., Straaton K. V., Moulds J. M., Lisse J., Bastian H. M., McGwin G., Jr., Bartolucci A. A., Roseman J. M., & Reveille J. D. 1999, "Systemic lupus erythematosus in three ethnic groups: III. A comparison of characteristics early in the natural history of the LUMINA cohort. LUpus in MInority populations: NAture vs. Nurture", *Lupus*, vol. 8, no. 3, pp. 197-209.

Alarcon G. S., McGwin G., Jr., Petri M., Reveille J. D., Ramsey-Goldman R., & Kimberly R. P. 2002, "Baseline characteristics of a multiethnic lupus cohort: PROFILE", *Lupus*, vol. 11, no. 2, pp. 95-101.

Alarcon-Segovia D., arcon-Riquelme M. E., Cardiel M. H., Caeiro F., Massardo L., Villa A. R., & Pons-Estel B. A. 2005, "Familial aggregation of systemic lupus erythematosus, rheumatoid arthritis, and other autoimmune diseases in 1,177 lupus patients from the GLADEL cohort", *Arthritis Rheum.*, vol. 52, no. 4, pp. 1138-1147.

Andrich, D. 1988, *Rasch models for measurement* Sage, Newbury Park, CA.

Aringer M., Graninger W. B., Steiner G., & Smolen J. S. 2004, "Safety and efficacy of tumor necrosis factor alpha blockade in systemic lupus erythematosus: an open-label study", *Arthritis Rheum.*, vol. 50, no. 10, pp. 3161-3169.

Austin H. A., III, Klippel J. H., Balow J. E., le Riche N. G., Steinberg A. D., Plotz P. H., & Decker J. L. 1986, "Therapy of lupus nephritis. Controlled trial of prednisone and cytotoxic drugs", *N.Engl.J.Med.*, vol. 314, no. 10, pp. 614-619.

Bae S. C., Koh H. K., Chang D. K., Kim M. H., Park J. K., & Kim S. Y. 2001, "Reliability and validity of systemic lupus activity measure-revised (SLAM-R) for measuring clinical disease activity in systemic lupus erythematosus", *Lupus*, vol. 10, no. 6, pp. 405-409.

Baechler E. C., Batliwalla F. M., Karypis G., Gaffney P. M., Ortmann W. A., Espe K. J., Shark K. B., Grande W. J., Hughes K. M., Kapur V., Gregersen P. K., & Behrens T. W. 2003,

"Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus", *Proc.Natl.Acad.Sci.U.S.A*, vol. 100, no. 5, pp. 2610-2615.

Bave U., Magnusson M., Eloranta M. L., Perers A., Alm G. V., & Ronnblom L. 2003, "Fc gamma RIIa is expressed on natural IFN-alpha-producing cells (plasmacytoid dendritic cells) and is required for the IFN-alpha production induced by apoptotic cells combined with lupus IgG", *J.Immunol.*, vol. 171, no. 6, pp. 3296-3302.

Bengtsson A. A., Sturfelt G., Truedsson L., Blomberg J., Alm G., Vallin H., & Ronnblom L. 2000, "Activation of type I interferon system in systemic lupus erythematosus correlates with disease activity but not with antiretroviral antibodies", *Lupus*, vol. 9, no. 9, pp. 664-671.

Blanco P., Palucka A. K., Gill M., Pascual V., & Banchereau J. 2001, "Induction of dendritic cell differentiation by IFN-alpha in systemic lupus erythematosus", *Science*, vol. 294, no. 5546, pp. 1540-1543.

Bombardier C., Gladman D. D., Urowitz M. B., Caron D., & Chang C. H. 1992, "Derivation of the SLEDAI. A disease activity index for lupus patients. The Committee on Prognosis Studies in SLE", *Arthritis Rheum.*, vol. 35, no. 6, pp. 630-640.

Brennan P. & Silman A. 1992, "Statistical methods for assessing observer variability in clinical measures", *BMJ*, vol. 304, no. 6840, pp. 1491-1494.

Brunner H. I., Feldman B. M., Bombardier C., & Silverman E. D. 1999, "Sensitivity of the Systemic Lupus Erythematosus Disease Activity Index, British Isles Lupus Assessment Group Index, and Systemic Lupus Activity Measure in the evaluation of clinical change in childhood-onset systemic lupus erythematosus", *Arthritis Rheum.*, vol. 42, no. 7, pp. 1354-1360.

Buyon J. P., Tamerius J., Belmont H. M., & Abramson S. B. 1992, "Assessment of disease activity and impending flare in patients with systemic lupus erythematosus. Comparison of the use of complement split products and conventional measurements of complement", *Arthritis Rheum.*, vol. 35, no. 9, pp. 1028-1037.

Casciola-Rosen L. A., Anhalt G., & Rosen A. 1994, "Autoantigens targeted in systemic lupus erythematosus are clustered in two populations of surface structures on apoptotic keratinocytes", *J.Exp.Med.*, vol. 179, no. 4, pp. 1317-1330.

Cervera R., Khamashta M. A., Font J., Sebastiani G. D., Gil A., Lavilla P., Mejia J. C., Aydintug A. O., Chwalinska-Sadowska H., de R. E., Fernandez-Nebro A., Galeazzi M., Valen M., Mathieu A., Houssiau F., Caro N., Alba P., Ramos-Casals M., Ingelmo M., & Hughes G. R. 2003, "Morbidity and mortality in systemic lupus erythematosus during a 10-year period: a comparison of early and late manifestations in a cohort of 1,000 patients", *Medicine (Baltimore)*, vol. 82, no. 5, pp. 299-308.

Chang E., Abrahamowicz M., Ferland D., & Fortin P. R. 2002a, "Comparison of the responsiveness of lupus disease activity measures to changes in systemic lupus erythematosus activity relevant to patients and physicians", *J.Clin.Epidemiol.*, vol. 55, no. 5, pp. 488-497.

Chang E. R., Abrahamowicz M., Ferland D., & Fortin P. R. 2002b, "Organ manifestations influence differently the responsiveness of 2 lupus disease activity measures, according to patients' or physicians' evaluations of recent lupus activity", *J.Rheumatol.*, vol. 29, no. 11, pp. 2350-2358.

Churg J, Bernstein J, & Glassock R J 1995, *Renal Disease: Classification and Atlas of Glomerular Diseases*, Second edn, Igaku-Shoin, New York, Tokyo.

Cohen J. 1960, "A coefficient of agreement for nominal scales", *Educational and Psychological Measurement*, vol. 20, pp. 37-46.

Cohen J. 1968, "Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit", *Psychological Bulletin*, vol. 70, pp. 213-220.

Cooper G. S., Parks C. G., Treadwell E. L., St Clair E. W., Gilkeson G. S., & Dooley M. A. 2004, "Occupational risk factors for the development of systemic lupus erythematosus", *J.Rheumatol.*, vol. 31, no. 10, pp. 1928-1933.

Cornacchia E., Golbus J., Maybaum J., Strahler J., Hanash S., & Richardson B. 1988, "Hydralazine and procainamide inhibit T cell DNA methylation and induce autoreactivity", *J.Immunol.*, vol. 140, no. 7, pp. 2197-2200.

Crow, M. K. & Kirou, K. A. 2007, "Cytokines and interferons in lupus," in *Dubois' Lupus Erythematosus*, Seventh Edition edn, D. J. Wallace & B. H. Hahn, eds., Lippincott Williams & Wilkins, Philadelphia, pp. 161-175.

D'Cruz D. 2001, "Antimalarial therapy: a panacea for mild lupus?", *Lupus*, vol. 10, no. 3, pp. 148-151.

Dall'Era M. & Wofsy D. 2006, "Clinical trial design in systemic lupus erythematosus", *Curr.Opin.Rheumatol.*, vol. 18, no. 5, pp. 476-480.

Davison, A. C. & Hinckley, D. V. 2007, *Bootstrap methods and their applications* Cambridge University Press, Cambridge.

Deapen D., Escalante A., Weinrib L., Horwitz D., Bachman B., Roy-Burman P., Walker A., & Mack T. M. 1992, "A revised estimate of twin concordance in systemic lupus erythematosus", *Arthritis Rheum.*, vol. 35, no. 3, pp. 311-318.

Enyedy E. J., Nambiar M. P., Liossis S. N., Dennis G., Kammer G. M., & Tsokos G. C. 2001, "Fc epsilon receptor type I gamma chain replaces the deficient T cell receptor zeta chain in T cells of patients with systemic lupus erythematosus", *Arthritis Rheum.*, vol. 44, no. 5, pp. 1114-1121.

Esdaile J. M., Abrahamowicz M., Joseph L., MacKenzie T., Li Y., & Danoff D. 1996, "Laboratory tests as predictors of disease exacerbations in systemic lupus erythematosus. Why some tests fail", *Arthritis Rheum.*, vol. 39, no. 3, pp. 370-378.

Fleiss J. L. & Cohen J. 1973, "The equivalence of weighted kappa and the intra class correlation coefficient as measures of reliability.", *Educational and Psychological Measurement*, vol. 33, pp. 613-619.

Fortin P. R., Abrahamowicz M., Clarke A. E., Neville C., Du B. R., Fraenkel L., & Liang M. H. 2000, "Do lupus disease activity measures detect clinically important change?", *J.Rheumatol.*, vol. 27, no. 6, pp. 1421-1428.

Fruchter O. & Edoute Y. 2005, "First presentation of systemic lupus erythematosus following ultraviolet radiation exposure in an artificial tanning device", *Rheumatology.(Oxford)*, vol. 44, no. 4, pp. 558-559.

Gershov D., Kim S., Brot N., & Elkon K. B. 2000, "C-Reactive protein binds to apoptotic cells, protects the cells from assembly of the terminal complement components, and sustains an antiinflammatory innate immune response: implications for systemic autoimmunity", *J.Exp.Med.*, vol. 192, no. 9, pp. 1353-1364.

Ginzler E. M., Dooley M. A., Aranow C., Kim M. Y., Buyon J., Merrill J. T., Petri M., Gilkeson G. S., Wallace D. J., Weisman M. H., & Appel G. B. 2005, "Mycophenolate mofetil or intravenous cyclophosphamide for lupus nephritis", *N.Engl.J.Med.*, vol. 353, no. 21, pp. 2219-2228.

Gladman D., Ginzler E., Goldsmith C., Fortin P., Liang M., Urowitz M., Bacon P., Bombardieri S., Hanly J., Hay E., Isenberg D., Jones J., Kalunian K., Maddison P., Nived O., Petri M., Richter M., Sanchez-Guerrero J., Snaith M., Sturfelt G., Symmons D., & Zoma A. 1996, "The development and initial validation of the Systemic Lupus International Collaborating Clinics/American College of Rheumatology damage index for systemic lupus erythematosus", *Arthritis Rheum.*, vol. 39, no. 3, pp. 363-369.

Gladman D. D., Goldsmith C. H., Urowitz M. B., Bacon P., Bombardier C., Isenberg D., Kalunian K., Liang M. H., Maddison P., Nived O., Richter M., Snaith M., Symmons D., & Zoma A. 1992, "Crosscultural validation and reliability of 3 disease activity indices in systemic lupus erythematosus", *J.Rheumatol.*, vol. 19, no. 4, pp. 608-611.

Gladman D. D., Goldsmith C. H., Urowitz M. B., Bacon P., Bombardier C., Isenberg D., Kalunian K., Liang M. H., Maddison P., Nived O., Richter M., Snaith M., Symmons D., & Zoma A. 1994, "Sensitivity to change of 3 Systemic Lupus Erythematosus Disease Activity Indices: international validation", *J.Rheumatol.*, vol. 21, no. 8, pp. 1468-1471.

Gladman D. D., Goldsmith C. H., Urowitz M. B., Bacon P., Fortin P., Ginzler E., Gordon C., Hanly J. G., Isenberg D. A., Petri M., Nived O., Snaith M., & Sturfelt G. 2000a, "The Systemic Lupus International Collaborating Clinics/American College of Rheumatology (SLICC/ACR) Damage Index for Systemic Lupus Erythematosus International Comparison", *J.Rheumatol.*, vol. 27, no. 2, pp. 373-376.

Gladman D. D., Hirani N., Ibanez D., & Urowitz M. B. 2003, "Clinically active serologically quiescent systemic lupus erythematosus", *J.Rheumatol.*, vol. 30, no. 9, pp. 1960-1962.

Gladman D. D., Ibanez D., & Urowitz M. B. 2002, "Systemic lupus erythematosus disease activity index 2000", *J.Rheumatol.*, vol. 29, no. 2, pp. 288-291.

Gladman D. D., Urowitz M. B., Goldsmith C. H., Fortin P., Ginzler E., Gordon C., Hanly J. G., Isenberg D. A., Kalunian K., Nived O., Petri M., Sanchez-Guerrero J., Snaith M., & Sturfelt G. 1997, "The reliability of the Systemic Lupus International Collaborating Clinics/American College of Rheumatology Damage Index in patients with systemic lupus erythematosus", *Arthritis Rheum.*, vol. 40, no. 5, pp. 809-813.

Gladman D. D., Urowitz M. B., Kagal A., & Hallett D. 2000b, "Accurately describing changes in disease activity in Systemic Lupus Erythematosus", *J.Rheumatol.*, vol. 27, no. 2, pp. 377-379.

Graham R. R., Ortmann W. A., Langefeld C. D., Jawaheer D., Selby S. A., Rodine P. R., Baechler E. C., Rohlf K. E., Shark K. B., Espe K. J., Green L. E., Nair R. P., Stuart P. E., Elder J. T., King R. A., Moser K. L., Gaffney P. M., Bugawan T. L., Erlich H. A., Rich S. S., Gregersen P. K., & Behrens T. W. 2002, "Visualizing human leukocyte antigen class II risk haplotypes in human systemic lupus erythematosus", *Am.J.Hum.Genet.*, vol. 71, no. 3, pp. 543-553.

Guzman J., Cardiel M. H., rce-Salinas A., Sanchez-Guerrero J., & Alarcon-Segovia D. 1992, "Measurement of disease activity in systemic lupus erythematosus. Prospective validation of 3 clinical indices", *J.Rheumatol.*, vol. 19, no. 10, pp. 1551-1558.

Hallegua D. S. & Wallace D. J. 2000, "Gastrointestinal manifestations of systemic lupus erythematosus", *Curr.Opin.Rheumatol.*, vol. 12, no. 5, pp. 379-385.

Hawker G., Gabriel S., Bombardier C., Goldsmith C., Caron D., & Gladman D. 1993, "A reliability study of SLEDAI: a disease activity index for systemic lupus erythematosus", *J.Rheumatol.*, vol. 20, no. 4, pp. 657-660.

Hay E. M., Bacon P. A., Gordon C., Isenberg D. A., Maddison P., Snaith M. L., Symmons D. P., Viner N., & Zoma A. 1993, "The BILAG index: a reliable and valid instrument for measuring clinical disease activity in systemic lupus erythematosus", *Q.J.Med.*, vol. 86, no. 7, pp. 447-458.

He B., Qiao X., & Cerutti A. 2004, "CpG DNA induces IgG class switch DNA recombination by activating human B cells through an innate pathway that requires TLR9 and cooperates with IL-10", *J.Immunol.*, vol. 173, no. 7, pp. 4479-4491.

Herrera-Esparza R., Barbosa-Cisneros O., Villalobos-Hurtado R., & valos-Diaz E. 1998, "Renal expression of IL-6 and TNFalpha genes in lupus nephritis", *Lupus*, vol. 7, no. 3, pp. 154-158.

Herzinger T., Plewig G., & Rocken M. 2004, "Use of sunscreens to protect against ultraviolet-induced lupus erythematosus", *Arthritis Rheum.*, vol. 50, no. 9, pp. 3045-3046.

Hirohata S. & Miyamoto T. 1990, "Elevated levels of interleukin-6 in cerebrospinal fluid from patients with systemic lupus erythematosus and central nervous system involvement", *Arthritis Rheum.*, vol. 33, no. 5, pp. 644-649.

Ho A., Barr S. G., Magder L. S., & Petri M. 2001a, "A decrease in complement is associated with increased renal and hematologic activity in patients with systemic lupus erythematosus", *Arthritis Rheum.*, vol. 44, no. 10, pp. 2350-2357.

Ho A., Magder L. S., Barr S. G., & Petri M. 2001b, "Decreases in anti-double-stranded DNA levels are associated with concurrent flares in patients with systemic lupus erythematosus", *Arthritis Rheum.*, vol. 44, no. 10, pp. 2342-2349.

Hochberg M. C. 1997, "Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus", *Arthritis Rheum.*, vol. 40, no. 9, p. 1725.

Hoffman R. W. 2004, "T cells in the pathogenesis of systemic lupus erythematosus", *Clin.Immunol.*, vol. 113, no. 1, pp. 4-13.

Hoffman R. W., Takeda Y., Sharp G. C., Lee D. R., Hill D. L., Kaneoka H., & Caldwell C. W. 1993, "Human T cell clones reactive against U-small nuclear ribonucleoprotein autoantigens from connective tissue disease patients and healthy individuals", *J.Immunol.*, vol. 151, no. 11, pp. 6460-6469.

Horii Y., Iwano M., Hirata E., Shiiki M., Fujii Y., Dohi K., & Ishikawa H. 1993, "Role of interleukin-6 in the progression of mesangial proliferative glomerulonephritis", *Kidney Int.Suppl*, vol. 39, p. S71-S75.

Houssiau F. A., Lefebvre C., Vanden B. M., Lambert M., Devogelaer J. P., & Renauld J. C. 1995, "Serum interleukin 10 titers in systemic lupus erythematosus reflect disease activity", *Lupus*, vol. 4, no. 5, pp. 393-395.

Husted J. A., Cook R. J., Farewell V. T., & Gladman D. D. 2000, "Methods for assessing responsiveness: a critical review and recommendations", *J.Clin.Epidemiol.*, vol. 53, no. 5, pp. 459-468.

Illei G. G., Austin H. A., Crane M., Collins L., Gourley M. F., Yarboro C. H., Vaughan E. M., Kuroiwa T., Danning C. L., Steinberg A. D., Klippel J. H., Balow J. E., & Boumpas D. T. 2001, "Combination therapy with pulse cyclophosphamide plus pulse methylprednisolone improves long-term renal outcome without adding toxicity in patients with lupus nephritis", *Ann.Intern.Med.*, vol. 135, no. 4, pp. 248-257.

Illei G. G., Tackey E., Lapteva L., & Lipsky P. E. 2004a, "Biomarkers in systemic lupus erythematosus. I. General overview of biomarkers and their applicability", *Arthritis Rheum.*, vol. 50, no. 6, pp. 1709-1720.

Illei G. G., Tackey E., Lapteva L., & Lipsky P. E. 2004b, "Biomarkers in systemic lupus erythematosus: II. Markers of disease activity", *Arthritis Rheum.*, vol. 50, no. 7, pp. 2048-2065.

Ioannou Y. & Isenberg D. A. 2000, "Current evidence for the induction of autoimmune rheumatic manifestations by cytokine therapy", *Arthritis Rheum.*, vol. 43, no. 7, pp. 1431-1442.

Isenberg D. & Ramsey-Goldman R. 1999, "Assessing patients with lupus: towards a drug responder index", *Rheumatology.(Oxford)*, vol. 38, no. 11, pp. 1045-1049.

Isenberg D. A. 2006, "B cell targeted therapies in autoimmune diseases", *J.Rheumatol.Suppl*, vol. 77, pp. 24-28.

Isenberg D. A. & Gordon C. 2000, "From BILAG to BLIPS--disease activity assessment in lupus past, present and future", *Lupus*, vol. 9, no. 9, pp. 651-654.

Isenberg D. A., Rahman A., Allen E., Farewell V., Akil M., Bruce I. N., D'Cruz D., Griffiths B., Khamashta M., Maddison P., McHugh N., Snaith M., Teh L. S., Yee C. S., Zoma A., & Gordon C. 2005, "BILAG 2004. Development and initial validation of an updated version of the British Isles Lupus Assessment Group's disease activity index for patients with systemic lupus erythematosus", *Rheumatology.(Oxford)*, vol. 44, no. 7, pp. 902-906.

Jacobi A. M., Odendahl M., Reiter K., Bruns A., Burmester G. R., Radbruch A., Valet G., Lipsky P. E., & Dorner T. 2003, "Correlation between circulating CD27high plasma cells and disease activity in patients with systemic lupus erythematosus", *Arthritis Rheum.*, vol. 48, no. 5, pp. 1332-1342.

Jacobson D. L., Gange S. J., Rose N. R., & Graham N. M. 1997, "Epidemiology and estimated population burden of selected autoimmune diseases in the United States", *Clin.Immunol.Immunopathol.*, vol. 84, no. 3, pp. 223-243.

James J. A., Kaufman K. M., Farris A. D., Taylor-Albert E., Lehman T. J., & Harley J. B. 1997, "An increased prevalence of Epstein-Barr virus infection in young patients suggests a possible etiology for systemic lupus erythematosus", *J.Clin.Invest*, vol. 100, no. 12, pp. 3019-3026.

James J. A., Neas B. R., Moser K. L., Hall T., Bruner G. R., Sestak A. L., & Harley J. B. 2001, "Systemic lupus erythematosus in adults is associated with previous Epstein-Barr virus exposure", *Arthritis Rheum.*, vol. 44, no. 5, pp. 1122-1126.

Johnson A., Harkin S., Steward M. W., & Whaley K. 1987, "The effects of immunoglobulin isotype and antibody affinity on complement-mediated inhibition of immune precipitation and solubilization", *Mol.Immunol.*, vol. 24, no. 11, pp. 1211-1217.

Johnson A. E., Cavalcanti F. S., Gordon C., Nived O., Palmer R. G., Sturfelt G., Viner N. J., & Bacon P. A. 1994, "Cross-sectional analysis of the differences between patients with systemic lupus erythematosus in England, Brazil and Sweden", *Lupus*, vol. 3, no. 6, pp. 501-506.

Johnson A. E., Gordon C., Palmer R. G., & Bacon P. A. 1995, "The prevalence and incidence of systemic lupus erythematosus in Birmingham, England. Relationship to ethnicity and country of birth", *Arthritis Rheum.*, vol. 38, no. 4, pp. 551-558.

Jury E. C., Kabouridis P. S., Flores-Borja F., Mageed R. A., & Isenberg D. A. 2004, "Altered lipid raft-associated signaling and ganglioside expression in T lymphocytes from patients with systemic lupus erythematosus", *J.Clin.Invest*, vol. 113, no. 8, pp. 1176-1187.

Karassa F. B., Bijl M., Davies K. A., Kallenberg C. G., Khamashta M. A., Manger K., Michel M., Piette J. C., Salmon J. E., Song Y. W., Tsuchiya N., Yoo D. H., & Ioannidis J. P. 2003a, "Role of the Fcgamma receptor IIA polymorphism in the antiphospholipid syndrome: an international meta-analysis", *Arthritis Rheum.*, vol. 48, no. 7, pp. 1930-1938.

Karassa F. B., Trikalinos T. A., & Ioannidis J. P. 2003b, "The Fc gamma RIIIA-F158 allele is a risk factor for the development of lupus nephritis: a meta-analysis", *Kidney Int.*, vol. 63, no. 4, pp. 1475-1482.

Kirou K. A., Lee C., George S., Louca K., Peterson M. G., & Crow M. K. 2005, "Activation of the interferon-alpha pathway identifies a subgroup of systemic lupus erythematosus patients with distinct serologic features and active disease", *Arthritis Rheum.*, vol. 52, no. 5, pp. 1491-1503.

Landis J. R. & Koch G. G. 1977, "The measurement of observer agreement for categorical data", *Biometrics*, vol. 33, no. 1, pp. 159-174.

Lang T. J. 2004, "Estrogen as an immunomodulator", *Clin.Immunol.*, vol. 113, no. 3, pp. 224-230.

Lee C. K., Ahn M. S., Lee E. Y., Shin J. H., Cho Y. S., Ha H. K., Yoo B., & Moon H. B. 2002, "Acute abdominal pain in systemic lupus erythematosus: focus on lupus enteritis (gastrointestinal vasculitis)", *Ann.Rheum.Dis.*, vol. 61, no. 6, pp. 547-550.

Lee Y. H., Harley J. B., & Nath S. K. 2005a, "CTLA-4 polymorphisms and systemic lupus erythematosus (SLE): a meta-analysis", *Hum.Genet.*, vol. 116, no. 5, pp. 361-367.

Lee Y. H., Witte T., Momot T., Schmidt R. E., Kaufman K. M., Harley J. B., & Sestak A. L. 2005b, "The mannose-binding lectin gene polymorphisms and systemic lupus erythematosus: two case-control studies and a meta-analysis", *Arthritis Rheum.*, vol. 52, no. 12, pp. 3966-3974.

Leong K. P., Kong K. O., Thong B. Y., Koh E. T., Lian T. Y., Teh C. L., Cheng Y. K., Chng H. H., Badsha H., Law W. G., Lau T. C., Chew L. C., Ho H. J., Pong L. Y., Hoi L. S., Sangeetha N., Chan S. P., & Howe H. S. 2005, "Development and preliminary validation of a systemic lupus erythematosus-specific quality-of-life instrument (SLEQOL)", *Rheumatology.(Oxford)*, vol. 44, no. 10, pp. 1267-1276.

Levin R. E., Weinstein A., Peterson M., Testa M. A., & Rothfield N. F. 1984, "A comparison of the sensitivity of the 1971 and 1982 American Rheumatism Association criteria for the classification of systemic lupus erythematosus", *Arthritis Rheum.*, vol. 27, no. 5, pp. 530-538.

Liang M., Fortin P., Schneider M., Abrahamowicz M., Alarcon G. S., Bombardieri S., Balow J., Benito-Garcia E., Bischoff-Ferrari H., Buyon J., Chehab G., Costenbader K., Crofford L., de Pablo P., Esdaile J. M., Finckh A., Fischer-Betz R., Gladman D., Gordon C., Illei G., Isenberg D., Iking-Konert C., Johnson K., Kalden J., Khamashta M., Koike T., Lockshin M., Manzi S., McCune J., Mikdashi J., Moore A., Mosca M., Petri M., Phillips C., Roberts N., Schur P., Smolen J., St Clair W., & Vibeke S. 2004, "The American College of

Rheumatology response criteria for systemic lupus erythematosus clinical trials: measures of overall disease activity", *Arthritis Rheum.*, vol. 50, no. 11, pp. 3418-3426.

Liang M. H., Socher S. A., Larson M. G., & Schur P. H. 1989, "Reliability and validity of six systems for the clinical assessment of disease activity in systemic lupus erythematosus", *Arthritis Rheum.*, vol. 32, no. 9, pp. 1107-1118.

Liang M. H., Socher S. A., Roberts W. N., & Esdaile J. M. 1988, "Measurement of systemic lupus erythematosus activity in clinical research", *Arthritis Rheum.*, vol. 31, no. 7, pp. 817-825.

Liu C. C., Manzi S., Kao A. H., Navratil J. S., Ruffing M. J., & Ahearn J. M. 2005, "Reticulocytes bearing C4d as biomarkers of disease activity for systemic lupus erythematosus", *Arthritis Rheum.*, vol. 52, no. 10, pp. 3087-3099.

Llorente L., Richaud-Patin Y., Fior R., cocer-Varela J., Wijdenes J., Fourrier B. M., Galanaud P., & Emilie D. 1994, "In vivo production of interleukin-10 by non-T cells in rheumatoid arthritis, Sjogren's syndrome, and systemic lupus erythematosus. A potential mechanism of B lymphocyte hyperactivity and autoimmunity", *Arthritis Rheum.*, vol. 37, no. 11, pp. 1647-1655.

Llorente L., Richaud-Patin Y., Garcia-Padilla C., Claret E., Jakez-Ocampo J., Cardiel M. H., cocer-Varela J., Grangeot-Keros L., arcon-Segovia D., Wijdenes J., Galanaud P., & Emilie D. 2000, "Clinical and biologic effects of anti-interleukin-10 monoclonal antibody administration in systemic lupus erythematosus", *Arthritis Rheum.*, vol. 43, no. 8, pp. 1790-1800.

Lobatto S., Daha M. R., Breedveld F. C., Pauwels E. K., Evers-Schouten J. H., Voetman A. A., Cats A., & Van Es L. A. 1988, "Abnormal clearance of soluble aggregates of human immunoglobulin G in patients with systemic lupus erythematosus", *Clin.Exp.Immunol.*, vol. 72, no. 1, pp. 55-59.

Lovgren T., Eloranta M. L., Bave U., Alm G. V., & Ronnblom L. 2004, "Induction of interferon-alpha production in plasmacytoid dendritic cells by immune complexes containing nucleic acid released by necrotic or late apoptotic cells and lupus IgG", *Arthritis Rheum.*, vol. 50, no. 6, pp. 1861-1872.

Maclure M. & Willett W. C. 1987, "Misinterpretation and misuse of the kappa statistic", *Am.J.Epidemiol.*, vol. 126, no. 2, pp. 161-169.

Manzi S., Navratil J. S., Ruffing M. J., Liu C. C., Danchenko N., Nilson S. E., Krishnaswami S., King D. E., Kao A. H., & Ahearn J. M. 2004, "Measurement of erythrocyte C4d and complement receptor 1 in systemic lupus erythematosus", *Arthritis Rheum.*, vol. 50, no. 11, pp. 3596-3604.

McCarty D. J., Manzi S., Medsger T. A., Jr., Ramsey-Goldman R., LaPorte R. E., & Kwoh C. K. 1995, "Incidence of systemic lupus erythematosus. Race and gender differences", *Arthritis Rheum.*, vol. 38, no. 9, pp. 1260-1270.

McElhone K., Abbott J., Shelmerdine J., Bruce I. N., Ahmad Y., Gordon C., Peers K., Isenberg D., Ferenkeh-Koroma A., Griffiths B., Akil M., Maddison P., & Teh L. S. 2007,

"Development and validation of a disease-specific health-related quality of life measure, the LupusQol, for adults with systemic lupus erythematosus", *Arthritis Rheum.*, vol. 57, no. 6, pp. 972-979.

MERRELL M. & SHULMAN L. E. 1955, "Determination of prognosis in chronic disease, illustrated by systemic lupus erythematosus", *J.Chronic.Dis.*, vol. 1, no. 1, pp. 12-32.

Mollnes T. E., Haga H. J., Brun J. G., Nielsen E. W., Sjoholm A., Sturfeldt G., Martensson U., Bergh K., & Rekvig O. P. 1999, "Complement activation in patients with systemic lupus erythematosus without nephritis", *Rheumatology.(Oxford)*, vol. 38, no. 10, pp. 933-940.

Moss K. E., Ioannou Y., Sultan S. M., Haq I., & Isenberg D. A. 2002, "Outcome of a cohort of 300 patients with systemic lupus erythematosus attending a dedicated clinic for over two decades", *Ann.Rheum.Dis.*, vol. 61, no. 5, pp. 409-413.

Mukhtyar C., Boumpas D., Gordon C., Gross W., Jayne D., & Luqmani R. 2007, "Why we need guidelines for clinical trials in vasculitis and systemic lupus erythematosus", *Ann.Rheum.Dis.*, vol. 66, no. 5, pp. 569-570.

Ng K. P., Manson J. J., Rahman A., & Isenberg D. A. 2006, "Association of antinucleosome antibodies with disease flare in serologically active clinically quiescent patients with systemic lupus erythematosus", *Arthritis Rheum.*, vol. 55, no. 6, pp. 900-904.

Norsworthy P., Theodoridis E., Botto M., Athanassiou P., Beynon H., Gordon C., Isenberg D., Walport M. J., & Davies K. A. 1999, "Overrepresentation of the Fcgamma receptor type IIA R131/R131 genotype in caucasoid systemic lupus erythematosus patients with autoantibodies to C1q and glomerulonephritis", *Arthritis Rheum.*, vol. 42, no. 9, pp. 1828-1832.

Nunnally J. C. & Bernstein I. H. Psychometric theory. 3rd Edition. 1994. New York, McGraw-Hill.
Ref Type: Generic

Nurnberg W., Haas N., Schadendorf D., & Czarnetzki B. M. 1995, "Interleukin-6 expression in the skin of patients with lupus erythematosus", *Exp.Dermatol.*, vol. 4, no. 1, pp. 52-57.

Parks C. G., Cooper G. S., Nylander-French L. A., Sanderson W. T., Dement J. M., Cohen P. L., Dooley M. A., Treadwell E. L., St Clair E. W., Gilkeson G. S., Hoppin J. A., & Savitz D. A. 2002, "Occupational exposure to crystalline silica and risk of systemic lupus erythematosus: a population-based, case-control study in the southeastern United States", *Arthritis Rheum.*, vol. 46, no. 7, pp. 1840-1850.

Petri M., Genovese M., Engle E., & Hochberg M. 1991, "Definition, incidence, and clinical description of flare in systemic lupus erythematosus. A prospective cohort study", *Arthritis Rheum.*, vol. 34, no. 8, pp. 937-944.

Petri M., Hellmann D., & Hochberg M. 1992, "Validity and reliability of lupus activity measures in the routine clinic setting", *J.Rheumatol.*, vol. 19, no. 1, pp. 53-59.

Petri M., Kim M. Y., Kalunian K. C., Grossman J., Hahn B. H., Sammaritano L. R., Lockshin M., Merrill J. T., Belmont H. M., Askanase A. D., McCune W. J., Hearth-Holmes M., Dooley M. A., Von F. J., Friedman A., Tan M., Davis J., Cronin M., Diamond B., Mackay M., Sigler L., Fillius M., Rupel A., Licciardi F., & Buyon J. P. 2005, "Combined oral contraceptives in women with systemic lupus erythematosus", *N.Engl.J.Med.*, vol. 353, no. 24, pp. 2550-2558.

Petri M. & Magder L. 2004a, "Classification criteria for systemic lupus erythematosus: a review", *Lupus*, vol. 13, no. 11, pp. 829-837.

Petri M. A., Lahita R. G., Van Vollenhoven R. F., Merrill J. T., Schiff M., Ginzler E. M., Strand V., Kunz A., Gorelick K. J., & Schwartz K. E. 2002, "Effects of prasterone on corticosteroid requirements of women with systemic lupus erythematosus: a double-blind, randomized, placebo-controlled trial", *Arthritis Rheum.*, vol. 46, no. 7, pp. 1820-1829.

Petri M. A., Mease P. J., Merrill J. T., Lahita R. G., Iannini M. J., Yocum D. E., Ginzler E. M., Katz R. S., Gluck O. S., Genovese M. C., Van V. R., Kalunian K. C., Manzi S., Greenwald M. W., Buyon J. P., Olsen N. J., Schiff M. H., Kavanaugh A. F., Caldwell J. R., Ramsey-Goldman R., St C. E., Goldman A. L., Egan R. M., Polisson R. P., Moder K. G., Rothfield N. F., Spencer R. T., Hobbs K., Fessler B. J., Calabrese L. H., Moreland L. W., Cohen S. B., Quarles B. J., Strand V., Gurwith M., & Schwartz K. E. 2004b, "Effects of prasterone on disease activity and symptoms in women with active systemic lupus erythematosus", *Arthritis Rheum.*, vol. 50, no. 9, pp. 2858-2868.

Pons-Estel B. A., Catoggio L. J., Cardiel M. H., Soriano E. R., Gentiletti S., Villa A. R., Abadi I., Caeiro F., Alvarellos A., & arcon-Segovia D. 2004, "The GLADEL multinational Latin American prospective inception cohort of 1,214 patients with systemic lupus erythematosus: ethnic and disease heterogeneity among "Hispanics"", *Medicine (Baltimore)*, vol. 83, no. 1, pp. 1-17.

Prodeus A. P., Goerg S., Shen L. M., Pozdnyakova O. O., Chu L., Alicot E. M., Goodnow C. C., & Carroll M. C. 1998, "A critical role for complement in maintenance of self-tolerance", *Immunity.*, vol. 9, no. 5, pp. 721-731.

Prokunina L., Castillejo-Lopez C., Oberg F., Gunnarsson I., Berg L., Magnusson V., Brookes A. J., Tentler D., Kristjansdottir H., Grondal G., Bolstad A. I., Svenungsson E., Lundberg I., Sturfelt G., Jonssen A., Truedsson L., Lima G., cocer-Varela J., Jonsson R., Gyllensten U. B., Harley J. B., arcon-Segovia D., Steinsson K., & arcon-Riquelme M. E. 2002, "A regulatory polymorphism in PDCD1 is associated with susceptibility to systemic lupus erythematosus in humans", *Nat.Genet.*, vol. 32, no. 4, pp. 666-669.

Pugh-Bernard A. E. & Cambier J. C. 2006, "B cell receptor signaling in human systemic lupus erythematosus", *Curr.Opin.Rheumatol.*, vol. 18, no. 5, pp. 451-455.

Rasch, G. 1980, *Probabilistic models for some intelligence and attaintment tests (Expanded edition)* University of Chicago Press, Chicago.

Reefman E., Kuiper H., Limburg P. C., Kallenberg C. G., & Bijl M. 2007, "Type I interferons are involved in the development of UVB-induced inflammatory skin lesions in Systemic Lupus Erythematosus (SLE) patients", *Ann.Rheum.Dis.*

Richardson B., Powers D., Hooper F., Yung R. L., & O'Rourke K. 1994, "Lymphocyte function-associated antigen 1 overexpression and T cell autoreactivity", *Arthritis Rheum.*, vol. 37, no. 9, pp. 1363-1372.

Rovere P., Sabbadini M. G., Vallinoto C., Fascio U., Recigno M., Crosti M., Ricciardi-Castagnoli P., Balestrieri G., Tincani A., & Manfredi A. A. 1999, "Dendritic cell presentation of antigens from apoptotic cells in a proinflammatory context: role of opsonizing anti-beta2-glycoprotein I antibodies", *Arthritis Rheum.*, vol. 42, no. 7, pp. 1412-1420.

Ruiz-Irastorza G., Khamashta M. A., Gordon C., Lockshin M. D., Johns K. R., Sammaritano L., & Hughes G. R. 2004, "Measuring systemic lupus erythematosus activity during pregnancy: validation of the lupus activity index in pregnancy scale", *Arthritis Rheum.*, vol. 51, no. 1, pp. 78-82.

Ruiz-Irastorza G., Lima F., Alves J., Khamashta M. A., Simpson J., Hughes G. R., & Buchanan N. M. 1996, "Increased rate of lupus flare during pregnancy and the puerperium: a prospective study of 78 pregnancies", *Br.J.Rheumatol.*, vol. 35, no. 2, pp. 133-138.

Salmon M. & Gordon C. 1999, "The role of apoptosis in systemic lupus erythematosus", *Rheumatology.(Oxford)*, vol. 38, no. 12, pp. 1177-1183.

Sanchez-Guerrero J., Karlson E. W., Liang M. H., Hunter D. J., Speizer F. E., & Colditz G. A. 1997, "Past use of oral contraceptives and the risk of developing systemic lupus erythematosus", *Arthritis Rheum.*, vol. 40, no. 5, pp. 804-808.

Sanchez-Guerrero J., Liang M. H., Karlson E. W., Hunter D. J., & Colditz G. A. 1995, "Postmenopausal estrogen therapy and the risk for developing systemic lupus erythematosus", *Ann.Intern.Med.*, vol. 122, no. 6, pp. 430-433.

Schifferli J. A., Steiger G., Hauptmann G., Spaeth P. J., & Sjoholm A. G. 1985, "Formation of soluble immune complexes by complement in sera of patients with various hypocomplementemic states. Difference between inhibition of immune precipitation and solubilization", *J.Clin.Invest*, vol. 76, no. 6, pp. 2127-2133.

Serdula M. K. & Rhoads G. G. 1979, "Frequency of systemic lupus erythematosus in different ethnic groups in Hawaii", *Arthritis Rheum.*, vol. 22, no. 4, pp. 328-333.

Sherer Y., Gorstein A., Fritzler M. J., & Shoenfeld Y. 2004, "Autoantibody explosion in systemic lupus erythematosus: more than 100 different antibodies found in SLE patients", *Semin.Arthritis Rheum.*, vol. 34, no. 2, pp. 501-537.

Shrout P. E. & Fleiss J. L. 1979, "Intraclass correlations: uses in assessing rater reliability", *Psychological Bulletin*, vol. 86, no. 2, pp. 420-428.

Singh J. A., Solomon D. H., Dougados M., Felson D., Hawker G., Katz P., Paulus H., & Wallace C. 2006, "Development of classification and response criteria for rheumatic diseases", *Arthritis Rheum.*, vol. 55, no. 3, pp. 348-352.

Sivaraj R. R., Durrani O. M., Denniston A. K., Murray P. I., & Gordon C. 2007, "Ocular manifestations of systemic lupus erythematosus", *Rheumatology.(Oxford)*, vol. 46, no. 12, pp. 1757-1762.

Smith C. D. & Cyr M. 1988, "The history of lupus erythematosus. From Hippocrates to Osler", *Rheum.Dis.Clin.North Am.*, vol. 14, no. 1, pp. 1-14.

Spronk P. E., Ter Borg E. J., Limburg P. C., & Kallenberg C. G. 1992, "Plasma concentration of IL-6 in systemic lupus erythematosus; an indicator of disease activity?", *Clin.Exp.Immunol.*, vol. 90, no. 1, pp. 106-110.

Stoll T., Gordon C., Seifert B., Richardson K., Malik J., Bacon P. A., & Isenberg D. A. 1997, "Consistency and validity of patient administered assessment of quality of life by the MOS SF-36; its association with disease activity and damage in patients with systemic lupus erythematosus", *J.Rheumatol.*, vol. 24, no. 8, pp. 1608-1614.

Stoll T., Seifert B., & Isenberg D. A. 1996, "SLICC/ACR Damage Index is valid, and renal and pulmonary organ scores are predictors of severe outcome in patients with systemic lupus erythematosus", *Br.J.Rheumatol.*, vol. 35, no. 3, pp. 248-254.

Studnicka-Benke A., Steiner G., Petera P., & Smolen J. S. 1996, "Tumour necrosis factor alpha and its soluble receptors parallel clinical disease and autoimmune activity in systemic lupus erythematosus", *Br.J.Rheumatol.*, vol. 35, no. 11, pp. 1067-1074.

Sturfelt G. & Truedsson L. 2005, "Complement and its breakdown products in SLE", *Rheumatology.(Oxford)*, vol. 44, no. 10, pp. 1227-1232.

Sultan S. M., Ioannou Y., & Isenberg D. A. 1999, "A review of gastrointestinal manifestations of systemic lupus erythematosus", *Rheumatology.(Oxford)*, vol. 38, no. 10, pp. 917-932.

Swaak A. J., Groenwold J., Aarden L. A., Statius van Eps L. W., & Feltkamp E. W. 1982, "Prognostic value of anti-dsDNA in SLE", *Ann.Rheum.Dis.*, vol. 41, no. 4, pp. 388-395.

Swaak A. J., Groenwold J., & Bronsveld W. 1986, "Predictive value of complement profiles and anti-dsDNA in systemic lupus erythematosus", *Ann.Rheum.Dis.*, vol. 45, no. 5, pp. 359-366.

Symmons D. P., Coppock J. S., Bacon P. A., Bresnihan B., Isenberg D. A., Maddison P., McHugh N., Snaith M. L., & Zoma A. S. 1988, "Development and assessment of a computerized index of clinical disease activity in systemic lupus erythematosus. Members of the British Isles Lupus Assessment Group (BILAG)", *Q.J.Med.*, vol. 69, no. 259, pp. 927-937.

Tan E. M., Cohen A. S., Fries J. F., Masi A. T., McShane D. J., Rothfield N. F., Schaller J. G., Talal N., & Winchester R. J. 1982, "The 1982 revised criteria for the classification of systemic lupus erythematosus", *Arthritis Rheum.*, vol. 25, no. 11, pp. 1271-1277.

Ter Borg E. J., Horst G., Hummel E. J., Limburg P. C., & Kallenberg C. G. 1990, "Measurement of increases in anti-double-stranded DNA antibody levels as a predictor of disease exacerbation in systemic lupus erythematosus. A long-term, prospective study", *Arthritis Rheum.*, vol. 33, no. 5, pp. 634-643.

Uramoto K. M., Michet C. J., Jr., Thumboo J., Sunku J., O'Fallon W. M., & Gabriel S. E. 1999, "Trends in the incidence and mortality of systemic lupus erythematosus, 1950-1992", *Arthritis Rheum.*, vol. 42, no. 1, pp. 46-50.

Vitali C., Bencivelli W., Isenberg D. A., Smolen J. S., Snaith M. L., Sciuto M., Neri R., & Bombardieri S. 1992, "Disease activity in systemic lupus erythematosus: report of the Consensus Study Group of the European Workshop for Rheumatology Research. II. Identification of the variables indicative of disease activity and their use in the development of an activity score. The European Consensus Study Group for Disease Activity in SLE", *Clin.Exp.Rheumatol.*, vol. 10, no. 5, pp. 541-547.

Walport M. J. 2002, "Complement and systemic lupus erythematosus", *Arthritis Res.*, vol. 4 Suppl 3, p. S279-S293.

Walz LeBlanc B. A., Gladman D. D., & Urowitz M. B. 1994, "Serologically active clinically quiescent systemic lupus erythematosus--predictors of clinical flares", *J.Rheumatol.*, vol. 21, no. 12, pp. 2239-2241.

Ward M. M., Marx A. S., & Barry N. N. 2000, "Comparison of the validity and sensitivity to change of 5 activity indices in systemic lupus erythematosus", *J.Rheumatol.*, vol. 27, no. 3, pp. 664-670.

Weening J. J., D'Agati V. D., Schwartz M. M., Seshan S. V., Alpers C. E., Appel G. B., Balow J. E., Bruijn J. A., Cook T., Ferrario F., Fogo A. B., Ginzler E. M., Hebert L., Hill G., Hill P., Jennette J. C., Kong N. C., Lesavre P., Lockshin M., Looi L. M., Makino H., Moura L. A., & Nagata M. 2004, "The classification of glomerulonephritis in systemic lupus erythematosus revisited", *Kidney Int.*, vol. 65, no. 2, pp. 521-530.

Williams R. L. 2000, "A note on robust variance estimation for cluster-correlated data", *Biometrics*, vol. 56, no. 2, pp. 645-646.

Wollaston S. J., Farewell V. T., Isenberg D. A., Gordon C., Merrill J. T., Petri M. A., & Kalunian K. C. 2004, "Defining response in systemic lupus erythematosus: a study by the Systemic Lupus International Collaborating Clinics group", *J.Rheumatol.*, vol. 31, no. 12, pp. 2390-2394.

Yee C. S., Isenberg D. A., Prabu A., Sokoll K., Teh L. S., Rahman A., Bruce I. N., Griffiths B., Akil M., McHugh N., D'Cruz D. P., Khamashta M. A., Maddison P., Zoma A., & Gordon C. 2007a, "BILAG-2004 Index captures SLE disease activity better than SLEDAI-2000", *Ann.Rheum.Dis.*

Yee C. S., Isenberg D. A., Prabu A., Sokoll K., Teh L. S., Rahman A., Bruce I. N., Griffiths B., Akil M., McHugh N., D'Cruz D. P., Khamashta M. A., Maddison P., Zoma A., & Gordon C. 2007b, "BILAG-2004 Index captures SLE disease activity better than SLEDAI-2000", *Ann.Rheum.Dis.*

Yurasov S., Wardemann H., Hammersen J., Tsuiji M., Meffre E., Pascual V., & Nussenzweig M. C. 2005, "Defective B cell tolerance checkpoints in systemic lupus erythematosus", *J.Exp.Med.*, vol. 201, no. 5, pp. 703-711.

Zweig M. H. & Campbell G. 1993, "Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine", *Clin.Chem.*, vol. 39, no. 4, pp. 561-577.