Multi-organ System Rheumatological Disease: Statistical Analysis of Outcome Measures and Their Interrelationships.

Elizabeth Jane Allen

Doctor of Philosophy

University College London

April 2003

UMI Number: U602481

UMI U602481

Multi-organ System Rheumatological Disease: Statistical Analysis of Outcome Measures and Their Interrelationships.

Idiopathic inflammatory myopathies are usually regarded as a heterogeneous group of autoimmune rheumatic diseases. Dermatomyositis and polymyositis may affect children and adults and, although rare, are a major cause of disability. In order to assess the value of conventional and newer therapies, a core set of measures for assessing myositis outcomes are being developed.

This thesis reports on the design and analysis of two real patient exercises carried out to study proposed measures. An approach to the study of reliability and agreement is presented. Inference procedures for ratios of standard errors are developed.

The myostitis measures are based on previous work in systemic lupus erythematosus (lupus), a major autoimmune rheumatic disease. International attempts to define validated disease activity and damage indices to assess patients with lupus have provided a consistent way to assess the disease. However, its multiple clinical manifestations prove a great challenge to rheumatologists managing patients with lupus. There is a need to better understand predictors of disease activity in order to improve and standardize therapy and to prevent the development of chronic damage.

This thesis presents an analysis of a clinical database for patients with lupus. The aim is to develop approaches to examine the interrelationships between disease activity in the different organ systems. The database available for analysis consists of data collected on 440 patients over a period of 10 years.

The analysis is based on logistic regression methodology with outcomes defined at the times of clinic visits. The usefulness of separate logistic regressions with dynamic covariates for the analysis of multinomial panel data

1

is illustrated. The efficiency of the approach relative to modelling disease activity in continuous time is investigated.

Acknowledgements

Contents

5

List of Tables

9

11

12

14

17

List of Figures

19

Chapter 1

Introduction

The immune system is a complex network of cells and organs that has evolved to defend the body against attacks by "foreign" invaders. One of the remarkable things about the immune system is its ability to recognize many millions of distinctive non-self molecules, and to respond by producing molecules that can match and counteract each one of the non-self molecules. At the heart of the immune response is the ability to distinguish between self and non-self, however, sometimes the immune system's recognition apparatus breaks down, and the body begins to manufacture antibodies and T cells directed against the body's own cells and organs. When this happens inflammation and autoimmunity may result.

Since the identification of autoimmune rheumatic disease, the underlying immune processes have been the subjects of intense clinical and laboratory research.

Idiopathic inflammatory myopathies constitute a heterogeneous group of diseases of unknown aetiology that are characterized by chronic inflammation of muscle tissue, skin, and other organs, leading to weakness, paralysis and if untreated, long term disability. Dermatomyositis, polymyositis and inclusion

body myositis (myositis) may affect children and adults and, although rare, are now widely recognized as a major cause of disability.

The most commonly prescribed drugs used to control disease progression are steroids and other anti-inflammatory or immunosuppressive agents, all of which have potentially serious side effects. It seems certain that these drugs have improved the outcome of patients with inflammatory muscle disease; nevertheless no combination has the capacity to induce a full remission in every case. The many side effects of the drugs remind the physician of the problems involved in treating the disease, with under-treatment leading to increased muscle weakness and over-treatment often resulting in increased morbidity. In order to assess more accurately the potential advantages and disadvantages of conventional and newer therapies as they are introduced, there is an urgent need to have available validated and reliable tools that assess disease activity and damage in myositis. Two real patient exercises took place, where newly developed tools were used to assess disease activity and damage in 7 patients. We attempt here to measure not only how well these tools differentiate between patients, but also to assess the level of agreement between physicians rating the same patient using these tools.

Systemic lupus erythematosus is a major autoimmune rheumatic disease and patients with this condition are cared for by virtually all rheumatologists. Its multiple clinical manifestations do however prove a great challenge to rheumatologists managing these patients and there is a continuing need to better understand predictors of disease activity in order to improve and standardize therapy and to prevent the development of chronic damage. Systemic lupus erythematosus causes significant morbidity and increased mortality, particularly for women at the prime of life.

In the past fifteen years there have been several international attempts to

define validated disease activity and damage indices to assess patients with systemic lupus erythematosus. These attempts have brought considerable order to what had become a disorganized situation.

We attempt in this thesis to develop approaches to data analysis that enable questions regarding interrelationships between disease activity in different systems and predictors and correlates of disease activity to be answered.

The database available for analysis consists of data collected on 440 patients with systemic lupus erythematosus over a period of 10 years. It incorporates considerable demographic information, measures of disease activity and damage and detailed information on therapy and laboratory data.

In this thesis we begin, in the second chapter, by giving a description of the clinical features of both idiopathic inflammatory myopathies and systemic lupus erythematosus with a brief description of their management. Validated activity and damage scores for the assessment of systemic lupus erythematosus are also described. The third chapter reports on the design and analysis of two real patient exercises carried out to study newly developed measures for the assessment of disease activity and damage in idiopathic inflammatory myopathies. An approach to the study of reliability and agreement is presented and inference procedures for ratios of standard errors are developed. The fourth chapter outlines methodologies used in the analysis of repeated categorical outcomes and describes the approach taken to the analysis of the systemic lupus erythematosus database. The usefulness of separate logistic regressions with dynamic covariates for the analysis of multinomial panel data is illustrated. The fifth chapter presents, with discussion, the results of all analyses. Chapter 6 investigates the efficiency of the approach taken relative to modelling disease activity in continuous time. The final chapter is a discussion of the work done.

Chapter 2

Systemic Lupus Erythematosus and Myositis

2.1 Systemic Lupus Erythematosus

Systemic lupus erythematosus (lupus) is a multi-system disease found world-wide that affects men, women and children of all ages (Belmont). Lupus can affect any part of the body, however skin involvement, joint pain and kidney involvement are among the more common problems. The course of the disease can vary from mild episodic illness to a severe and possibly fatal disease. A patient's symptoms can also vary widely over time and are usually characterized by periods of remission and exacerbation. At the onset of lupus, only one organ and or system may be involved. However additional organs invariably become involved later.

Lupus is characterized by the presence in the blood of particular autoantibodies to DNA, RNA and a number of other nuclear, cytoplasmic or cell-surface antigens. During periods of disease activity, the autoantibodies react with the 'self' antigens and immune complexes start to form, often building

up, causing inflammation, injury to tissues, and pain in various parts of the body.

Given the diversity of the symptoms of lupus and because the symptoms overlap with those of many other conditions, lupus can be difficult to diagnose and can be overlooked. To help distinguish lupus from other diseases, in 1971 the American Rheumatism Association (now the American College of Rheumatology) published a list of abnormalities (revised (Tan *et al.* 1982))) which point to lupus. The nature of the disease means that patients may visit a number of different specialists, and physicians should therefore consider lupus when assessing a number of conditions.

Studies have shown that women are about 10 to 20 times more likely to have, or develop, lupus than men, and that the disease is most commonly diagnosed in women between the ages of 20 and 40. However it is reported that about 20% of lupus cases develop before the age of 18, and that about 10 to 15% of cases develop after the age of 50. Lupus occurs more often in boys than men with boys representing 20 to 40% of children with lupus (Morrow *et al.* 1999). Lupus is usually more severe in younger patients and milder in older.

It is widely recognized that people from different racial and ethnic backgrounds are at different risk of developing lupus. A study from Birmingham (Johnson *et al.* 1995) recorded UK prevalence rates of 36.2, 90.6, and 206 per 100 000 among women of Caucasian, Asian, and Afro-Caribbean origin respectively. Other studies have shown that people of Chinese and Polynesian backgrounds also have an increased risk of developing lupus compared to Caucasians (Samanta *et al.* 1991).

Until 25 years ago, lupus was considered to be an acute and often fatal disease. However advances in diagnosis and treatment have greatly improved

prognosis with current expected ten year survival rates being greater than 80% (Boumpas *et al.* (1995) and Barlow *et al.* (1987)). While the use of corticosteroids and immunosuppressives has helped increase this survival rate, these drugs are still a cause of morbidity in lupus and there is still a need for further clarification of the pathogenesis of the disease for the treatment to become more targeted leading to a reduction in morbidity (Rasaratnam and Ryan 1995).

2.1.1   Clinical Features

Non-Specific Features

Fatigue, one of the most common features of lupus, is experienced by approximately 90% of patients and frequently described as unnatural. Lupus patients may also suffer from weight loss, anorexia, and frequently present with swollen glands (lymphadenopathy). However, infections and malignancies may also enlarge lymph glands, and before lupus is treated these possibilities should be ruled out. Fever is associated with lupus and is most commonly a feature of a flare of the disease. These non-specific features contribute to the difficulty in diagnosing lupus.

Musculoskeletal Involvement

The musculoskeletal system is commonly affected. Surveys have suggested that 80 to 90 % of patients with lupus suffer from pain in the joints (arthralgias) with arthritis (visible inflammation in a joint) being present in less than half these cases (Morrow *et al.* 1999). The joints most commonly involved are the proximal, interphalangeal, metacarpophalangeal, wrist, and knees. In contrast to rheumatoid arthritis, however, deforming joint abnormalities are

observed in only about 5% of lupus patients. The most common symptoms of arthritis in lupus patients are stiffness and aching.

Approximately 50% of lupus patients complain of muscle aches (myalgias), generally thought to be related to arthritis in nearby joints (Morrow *et al.* 1999). Another less common cause of muscle pain in lupus is myositis that occurs occasionally during the course of lupus ($\approx 3 - 5\%$). Steroid induced myopathy is a potential source of confusion. However, with inflammatory muscle disease, there is usually an elevation of the muscle enzymes.

Dermatological Involvement

Approximately 80% of patients with lupus have dermatological manifestations during the course of their illness. Practically any type of rash can occur many of which are exacerbated by photosensitivity. Only about 35% of patients however report the classic butterfly rash, a red rash occurring across the bridge of the nose and on the cheeks. Approximately 25% of lupus patients have discoid skin lesions. These lesions are often on the face or the inner pinna of the ear. Inflammation of the superficial blood vessels, known as cutaneous vasculitis, is seen in up to 70% of lupus patients and, if untreated, can result in ulceration or breakdown of the skin.

Alopecia occurs in about 50% of patients. Typically this manifests as reversible hair thinning during periods of disease activity. Discoid lesions involving the scalp can sometimes however lead to scarring alopecia. Mucosal ulcers occur in approximately 30% of lupus patients; the sores may be solitary or appear as crops of lesions. Approximately 30% of lupus patients suffer from Raynaud's phenomenon and 20 to 30% of all lupus patients have a red mottling under the skin, livedo reticularis.

Unusual cutaneous manifestations of lupus include urticaria, angioedema,

26

bullae and panniculitis (lupus profundus).

Cardiovascular and Pulmonary Involvement

The commonest manifestations of lupus in the heart or lungs involve the linings of these organs; the pleura and the pericardium. Symptoms of heart and/or lung involvement include chest pain, shortness of breath and a cough.

Inflammation of the pleura (pleurisy) is a common though usually not serious side effect of lupus. The symptom is pain although more severe forms of inflammation cause pleural effusions (fluid) leading to shortness of breath. Interstitial lung disease that involves the supporting structures of the lungs and pulmonary emboli are two other more frequent lung complications. Acute lupus pneumonitis, pulmonary haemorrhage and pulmonary hypertension are serious conditions that are sometimes observed and are difficult to treat. In a small number of patients a scarring (fibrosis) occurs on the lungs (Haupt *et al.* 1981).

Lupus patients frequently complain of chest pain that may or may not be related to heart disease. The true sources of cardiac pain are generally due to the most common cardiac manifestation pericarditis (inflammation of the pericardium) which manifests itself both with and without effusions.

Approximately 10% of lupus patients present with myocardial disease that is often serious, and may include inflammation in the form of myocarditis or heart muscle dysfunction that sometimes leads to congestive heart failure. It is of interest to note that autopsies reveal that in fact 40% of patients with lupus show evidence of prior myocardial involvement. Chest X-rays frequently show an enlarged heart, and signs of congestive heart failure are sometimes evident.

The classic Libman-Sachs non-bacterial endocarditis produces vegetation

on the mitral and aortic valves. Although clinically manifested as a heart murmur, the vegetations are usually so small that they are detectable by an echocardiogram only 30% of the time and are frequently asymptomatic and an incidental pathologic finding at autopsy. Although the vegetation produced alters the dynamics of the heart only 1 to 2% of the time, there are serious complications that can arise. They are prone to infection that can lead to subacute bacterial endocarditis where the vegetation become a growth site for bacteria, and occasionally parts of the vegetation may flake off causing a cerebral clot or stroke.

There is an increased incidence of premature coronary artery disease in lupus patients (now one of the most common causes of death) and high blood pressure, both aggravated by chronic steroid therapy (Wallance (1995), Bruce *et al.* (2000)).

Gastrointestinal Involvement

Around 45% of lupus patients suffer from gastrointestinal problems including nausea, mild abdominal pain, and diarrhoea. Severe inflammation of the intestinal tract occurs in less than 5% of patients and rarely, intestinal perforation occurs and can be life threatening. Fluid retention and swelling can cause intestinal obstruction.

The liver can be affected as a result of the lupus itself, or by the medications used to treat inflammation in lupus. Jaundice is found in 1 to 4% of lupus patients and hepatomegaly (enlargement of the liver) is found in up to 10% of lupus patients. Liver function abnormalities in lupus, however are most commonly explained by idiosyncratic reactions to drug therapy. Progression to cirrhosis as a consequence of inflammatory liver disease in lupus is rare. Mackay *et al.* (1959), first identified a form of lupoid hepatitis or

autoimmune hepatitis.

Inflammation of the pancreas is rare but does occasionally occur (Watts and Isenberg 1989).

Haematological Abnormalities

Anaemia is a common feature of lupus with about 80% of lupus patients being anaemic during the course of their disease. In lupus patients, anaemia may be caused by a number of different factors; chronic inflammation, iron deficiency and/or chronic renal disease. About 10% of lupus patients develop haemolytic anaemia a condition where red blood cells are prematurely destroyed.

Around half of lupus patients develop a low white blood cell count during the course of their disease (lymphopaenia). If a patient's white blood cell count falls below 3000, they are thought to have leukopaenia and it may suggest that they are about to have a flare.

Idiopathic thrombocytopaenic purpura, a decreased platelet count along with the presence of platelet antibodies, may affect lupus patients. A rare but serious complication of lupus however is thrombotic thrombocytopaenia purpura that can lead to multiple organ failure (Wallance 1995).

Lupus patients can also have qualitative platelet defects. Functional blood clotting can be affected by aspirin, chronic renal failure and platelet antibodies. This is however a benign condition requiring no treatment if platelet counts are in the normal range.

Nervous System Involvement

The majority of patients with lupus will suffer from some form of central nervous system disease. The most common symptom being cognitive dysfunc-

tion characterized by confusion, memory impairment and difficulty in articulating thoughts. Headaches are another common feature of lupus, with lupus patients being perhaps twice as likely to suffer from migraine like headaches than the general population.

Depression is an important manifestation of lupus, being in some patients an integral part of the disease with no one specific cause (Shortall *et al.* 1995). The most common cause is generally thought to be difficulty in dealing with the continuous series of stresses and strains associated with coping with chronic illness. However the disease itself may cause the depression and may then be helped by the management of the lupus itself.

Lupus patients can experience a variety of psychiatric disorders with some being wrongly diagnosed as schizophrenic. However it appears that treatment of the lupus generally results in a total improvement in the psychiatric features.

Some lupus patients may suffer from seizures or a series of epileptic fits and occasionally patients develop chorea. It is thought that 10% of patients with lupus develop inflammation of the peripheral nervous system at some point (Morrow *et al.* 1999).

A rare but serious effect of lupus is lupus myelitis that can include paralysis or weakness.

Renal Involvement

The kidneys are potentially the organs whose involvement is the most serious in lupus. As in most kidney diseases there are few obvious symptoms with little pain and most often the only indication is an abnormal urine or blood test. The most common major kidney problem is the leakage of protein into the urine. This can be mild and detected only on testing, or severe

gradually leading to a lowering of the protein level in the blood. If the kidneys become inflamed, the patient's blood pressure frequently rises and blood pressure measurement is important in the examination of lupus patients. If the kidneys become severely damaged their normal filtering process will be impaired, leading to the build up of toxic elements such as urea and creatinine in the blood, which may in turn lead to weight loss and nausea.

More complicated cases of renal involvement, particularly where treatment is needed, usually require a biopsy to make the diagnosis, and may require further biopsies to determine how the disease is progressing. A World Health Organisation or a National Institutes of Health classification system can be used to categorize the pathology of the kidney. This helps physicians judge the chance of response to treatment from the results of the biopsy.

Because the kidneys are so important to overall health, lupus affecting the kidneys generally requires intensive drug treatment to prevent permanent damage. If the kidney damage reaches a stage where toxic chemicals build up then dialysis is vital. In the most severe cases renal transplantation is considered. It is of interest to note that it has generally been found that lupus does not return to the transplanted kidney and in general patients with lupus who have renal transplantation do very well.

2.1.2 Management of Systemic Lupus Erythematosus

There is no cure for lupus. The objective in managing a disease like lupus is to control the severity of the symptoms, and to avoid a flare. General advice given to the lupus patient includes avoiding the sun, exercising, adherence to medications, and regular health care visits.

Several types of drugs are used to treat lupus. The treatment the doctor chooses is based on the patient's individual symptoms and needs.

For patients with joint or chest pain or fever, nonsteroidal anti-inflammatory drugs (NSAIDs) are often used. The main drawback of these drugs is that they have a number of side effects including stomach upset, heartburn, diarrhoea, and fluid retention. Some lupus patients also develop more serious side effects while taking NSAIDs including liver and kidney inflammation.

Antimalarials are another type of drug commonly used to treat lupus. They may be used alone or in combination with other drugs and are generally used to treat fatigue, joint pain, skin rashes, and inflammation of the lungs. Side effects of antimalarials can include stomach upset and, extremely rarely, damage to the retina of the eye.

The mainstay of lupus treatment involves the use of corticosteroids. These are life saving in acute lupus and can be given by mouth, in creams applied to the skin, or by injection. Modern treatment is geared to seek the lowest dose with the greatest benefit. The side effects of corticosteroids include swelling, increased appetite, weight gain, muscle weakness and over a period of time osteoporosis.

For patients whose kidneys or central nervous systems are affected by lupus, immunosuppressives may be used. Immunosuppressives block the production of some immune cells and curb the action of others. Side effects may include nausea, vomiting, hair loss, bladder problems, decreased fertility, and increased risk of cancer and infection. The risk for side effects increases with the length of treatment.

Both dialysis and renal transplants have proved effective in patients with renal disease.

2.1.3 Overall Clinical Assessment of Systemic Lupus Erythematosus

Lupus is an extremely clinically diverse disease that presents a substantial challenge to both the physician and the clinical researcher. In order to thoroughly assess the effects of lupus it is necessary to distinguish between disease activity (reversible clinical features) and disease damage (non-reversible change not related to active inflammation) and to consider the patient's own perception of their health.

There are a number of systems that have been devised for the assessment of disease activity in lupus. Global score systems include the Systemic Lupus Activity Measure (SLAM) (Liang *et al.* 1988) and the Systemic Lupus Erythematosus Disease Activity Index (SLEDAI) (Bombardier *et al.* 1992). In contrast the British Isles Lupus Assessment Group system (BILAG) (Hay *et al.* 1994) rates lupus activity in eight organs and/or systems. These indices have been shown to be valid and reliable (Liang *et al.* 1989).

The BILAG Index is a comprehensive index based developed according to the principle of the 'physicians intention to treat'. The index allocates separate alphabetic scores to each of eight organs and/or systems using the following ratings: A is the most active disease state requiring major immunosuppressive drugs, B the patient is known to have active disease but is already on immunosuppressive therapy, C the patient has relatively mild disease controlled by little specific therapy if any, D there is no activity in this system now and E no evidence of activity now or previously (Hay *et al.* 1993). Thus it can be easily discerned not only that a lupus patient is flaring but also in which particular system. For the purposes of comparison the BILAG index can be converted into a global score (A grade = 9 points, B = 3, C = 1, D and E =0) (Morrow *et al.* 1999). Strong correlations have

been shown between a BILAG global score and other commonly used global measures notably the SLAM and SLEDAI (Gladman *et al.* 1992).

The SLICC/ACR damage index was developed by the Systemic Lupus International Collaborating Clinics and accepted by the ACR (American College of Rheumatology) as a valid measure of damage in lupus (Gladman *et al.* 1997). The index includes descriptors in 12 organs and or systems and damage is only considered if present for at least 6 months.

The SF36 health assessment questionnaire is used in the study of lupus patients. The SF36 was designed for use in clinical practice and research, health policy evaluations, and general population surveys. It includes multi-item scales that assess eight health concepts and a health rating item where patients indicate a change in their health over a one year period.

It has been shown (Stoll *et al.* 1997) that in order to capture the total effect of lupus on patients all three types of indices are needed.

2.2 Myositis

There are three major forms of myositis

• Polymyositis, a condition involving the muscles that usually responds to immunosuppressive therapy.

• Dermatomyositis, a condition affecting the skin and the muscles that also usually responds to immunosuppressive therapy.

• Inclusion body myositis, a condition affecting the muscles with inclusion (abnormal protein deposits) and vacuoles (holes) in the muscle cells and fibres. Inclusion body myositis responds poorly to immunosuppressive therapy.

It is estimated that annually five to seven people per million will develop

a form of myositis. The incidence of polymyositis/dermatomyositis has been reported as 0.5 cases per 100 000 (Medsger *et al.* 1970) with a prevalence of 8 per 100 000 (DeVere and Bradley 1975).

Although myositis can affect people of any age, most children who get the disease tend to be between 5 and 15 years of age with most adults being over the age of 30. Dermatomyositis is much more common than polymyositis in children and both polymyositis and dermatomyositis are more common in women than men. However inclusion body myositis affects more men than women. Medsger *et al.* (1970) noted that black females were more likely to develop both polymyositis and dermatomyositis than any other group.

Myositis affects the muscles and connective tissues of the body, with the main symptom being muscular weakness that is usually progressive and can be severely disabling. It generally affects the muscles closest to the trunk such as those in the hips, thighs, shoulders, upper arms and neck. Patients may have difficulty in performing daily activities such as standing up, getting out of a car, climbing stairs, working overhead or combing their hair. About 70% of patients with dermatomyositis suffer from a characteristic rash that covers the knuckles, elbows or knees. It consists of small purplish red areas that are either flat and smooth or slightly raised (papules of Gottron). In more than half the patients with dermatomyositis the eyelids and skin around the eyes can become puffy and develop a heliotrope (lilac) rash. Patients with dermatomyositis may suffer from itchy and flaky scalps and any skin that is exposed to sun can become red and blotchy. Occasionally the skin directly around the fingernails may show enlarged blood vessels (hyperaemia).

2.2.1 Clinical Features

Joint Manifestations

Arthralgias/Arthritis are frequently present in myositis. They often occur concurrently with, or precede, the onset of muscle weakness and usually occur early in the disease course. The symptoms are more often than not, mild.

Cardiac Manifestations

Recent studies have found signs of cardiac involvement in more than of 70% of patients with polymyositis/dermatomyositis (Taylor *et al.* 1993). The major manifestations include conduction disturbances, arrhythmias and myocarditis. Pericardial effusions are seen in 5 to 25% of polymyositis/dermatomyositis patients but are usually asymptomatic (Tami and Bhasin 1993).

Pulmonary Manifestations

Pulmonary involvement resulting from muscle weakness, treatment or the underlying disease occurs in 40 to 50% of patients (Dickey and Myer 1974). Short rapid breathing (dyspnoea) and a non-productive cough are common pulmonary symptoms. The respiratory muscles are not generally affected until late in the course of the disease, however weakness will then occur. This weakness when manifested concurrently with pharyngeal weakness may predispose the patient to interstitial pneumonia and pulmonary fibrosis.

Other Systemic Manifestations

If the muscles in the oesophagus are affected patients may suffer from dysphagia (difficulty in swallowing). This in turn may cause weight loss and

36

malnutrition. Fatigue is also common.

Myositis and Cancer

There appears to be an association between both polymyositis and dermatomyositis and cancer. It is generally assumed that dermatomyositis rather than polymyositis represents an increased risk of malignancy (Sigurgeirsson *et al.* 1992) however recent population-based cohort studies show a stronger association between polymyositis and malignancy than previously thought. It appears that ovarian and lung cancers are associated with dermatomyositis while lung cancer and non-Hodgkin's lymphoma are associated with polymyositis. An association between malignant disease and inclusion body myositis has also been established for the first time (Buchbinder and Hill 2002).

Polymyositis and Dermatomyositis in children

Juvenile idiopathic inflammatory myopathy or juvenile myositis most often presents itself as dermatomyositis with its typical rashes and muscle weakness. In juvenile myositis the skin rashes are particularly prominent preceding muscle weakness more than 50% of the time. Muscle weakness usually develops, and as in adults, generally affects the proximal muscles (those closest to and within the trunk of the body). The neck, hip, trunk and shoulder muscles are primarily involved, but the distal muscles may also be affected. Dysphagia (difficulty swallowing), dysphonia (hoarseness), abdominal pain and arthritis can also occur in children with myositis and muscle pain is found in approximately 50% of children.

Calcinosis can affect children who have had the disease for a long time. Calcium crystals are deposited in large amounts throughout the body and

can lead to skin ulcers that are difficult to heal.

2.2.2 Management of Myositis

Corticosteroids are usually the first choice in the treatment of myositis (Oddis
and Medsger 1989). If however the disease does not respond adequately to
corticosteroids, immunosuppressants, notably azathioprine and methotrex-
ate may also be considered (Joffe *et al.* 1993). When in combination, im-
munosuppressants can be used to lessen the dose and potential side effects
of the corticosteroids. Because of the side-effects of these drugs, new ideas
on treatment are being actively pursued. These new therapies include TNF
α blockade and B cell depletion. To date however no adequate trials have
been undertaken.

In the case of a persistent rash (dermatomyositis) antimalarials may be
prescribed.

Calcinosis is difficult to treat, however it is believed that early and inten-
sive treatment may decrease the risk of serious complications.

Physical therapy can help maintain and improve muscle strength and
flexibility.

2.2.3 Motivation for the Development of Tools for the Assessment of Disease Activity and Damage in Myositis

As with lupus (Stoll *et al.* 1997), indeed as with many diseases, it seems
generally agreed that to capture the effects of myositis on a patient it is nec-
essary to assess disease activity, disease damage and the patient's perception
of the disease (F.W.Miller *et al.* 2001).

38

It is generally felt that (F.W.Miller *et al.* 2001), despite the absence of double blind controlled trials, corticosteroids, immunosuppressive and other drugs have a beneficial effect on the outcome of patients with inflammatory disease. However there are many side effects associated with these drugs which can result in serious morbidity. In order to assess the advantages and disadvantages of current therapies, and of new therapies as they are introduced, there is a need to have available validated and reliable tools that assess disease activity and damage in myositis. Furthermore it is recognized that although the skeletal muscles 'bear the brunt' of the disease, other organs and systems (e.g. the skin, lungs, heart and gastrointestinal tract) may also be involved. Thus rather analogous to lupus, it is evident that assessment tools need to record activity in a multiplicity of organs and or systems.

Myositis is a rare disease and, in order to make any reasonable sense of it, it is crucial that all patients are assessed in the same way. It would therefore clearly be an advantage if all clinical research groups were using the same indices. This would be beneficial both for long term observational studies and for clinical trials.

There are no previously available measures for the assessment of disease activity or damage in myositis. The organisation of a major international consensus for the development of disease activity and damage indices for the assessment of myositis is described in chapter 3.

Chapter 3

Myositis

3.1 The Development of Tools for the Assessment of Disease Activity and Damage in Myositis

In June 1999, at the European Union League Against Rheumatism (EULAR) meeting, an international dialogue about the assessment of disease activity and damage in patients with Myositis took place. Subsequent to this discussion a group of adult and paediatric specialists, with expertise in myopathies, met in March 2000 to develop a core set of measures for the assessment of myositis outcomes. Further development and consensus on these measures was then achieved via Delphi methods[1] by a multi-disciplinary group of over 70 rheumatologists, neurologists, rehabilitation specialists and others who constitute the International Myositis Outcome Assessment Collabora-

---

[1]A method of forecasting events by analysing the results of questionnaire sent to a panel of experts, who are therefore not subject to the inhibiting factors of a round table discussion.

tive Study Group (IMACS). Not surprisingly there was considerable debate on the best way to proceed. In order to achieve maximal international collaboration the physicians involved developed different tools for activity and damage.

Once these measures had been developed a need for them to be tested and validated was identified. Consequently in collaboration with Professor D. Isenberg and IMACS the real patient exercises, described subsequently, were planned. The primary goal of the exercise was to assess the performance of the newly developed measures. However it became evident that an added goal should be to establish whether these measures would provide physicians with a relatively quick evaluation of myositis patients.

3.1.1   The Assessment of Disease Activity

Two tools were developed for the assessment of disease activity. The Myositis Intention to Treat Activity Index (MITAX) and the Myositis Activity Score (MYOACT).

The MITAX is essentially a modification of the BILAG index used for the assessment of disease activity in patients with lupus. It is based on the physician's intent to treat principle and disease activity is graded from the most active grade A (requiring major immunosuppression) to grade E (no evidence of disease activity currently or previously in an organ or system). Each of the items included in this index were carefully considered by the group. Individual clinical features or combinations of features that the group anticipated would lead to the prescription of large doses of corticosteroids and/or immunosuppressives define a grade A, the most active score in each organ or system. For those patients with known disease activity that is ongoing and that continues to require somewhat lower doses of immunosuppressives

41

and/or other drugs, sets of criteria were used to define a B grade. The C grade in each organ or system defines patients with mild persistent activity. The D grade implies that the organ or system was once active but is no longer and the E grade implies that the organ or system is not active now, and has never been in the past. As with the BILAG index, the MITAX can be converted into a score with an A grade = 9 points, a B grade = 3 points, a C grade = 1 point, a D grade = 0 points and an E grade=0 points.

The MYOACT (Myositis activity score) consists of a series of 100mm visual analogue scales (the higher the number the more active the disease) which are completed by the physician assessing the patients.

3.1.2   The Assessment of Disease Damage

Two tools were developed for the assessment of disease damage, the Myositis Damage Index (MDI) and the Myositis Damage score (MYODAM). Damage implies persistent/permanent change in anatomy, physiology, pathology or function that is considered if it occurs after the diagnosis of the disease. No attempt is made in these tools, as with the SLICC Damage Index for patients with lupus, to ascribe the particular cause of the damage. The utility of a relatively simple damage index has been demonstrated in recent studies of patients with lupus in whom the early acquisition of damage in the first couple of years following diagnosis is a very powerful diagnostic marker for prognosis (Nived *et al.* 2000).

The Myositis Damage Index (MDI) is an assessment of damage developing in different organs and/or systems that is essentially a modification of the SLICC/ACR damage index. Each item assessed is given a score of either 0 (no damage) or 1 (damage).

The MYODAM (Myositis damage score) consists of a series of 100mm

visual analogue scales which are completed by the physicians and used to assess damage in the same organs and/or systems as the MDI.

3.2 Two Real Patient Exercises

In March 2001 in collaboration with Professor Isenberg an exercise to assess the performance of the newly developed tools was set up. In order to ensure that the measures provided physicians with a relatively quick assessment of the patients the exercise was run in one day. Seven patients with Myositis (4 female, 3 male, 4 with polymyositis, 3 with dermatomyositis) were assessed by 7 different physicians (6 rheumatologists, 1 neurologist). The physicians were provided with a one page synopsis of each patient's history and were asked to complete both of the activity and both of the damage scores. The physicians were also asked to complete a formal assessment of muscle strength using manual muscle tests. In particular they were asked to record strength in the neck flexors, deltoid, biceps, wrist extensors, gluteus maximus, gluteus medius, quadriceps and ankle dorsiflexors. Each consultation took one hour.

All forms of assessment were demonstrated prior to the exercise on patients with active disease. (The patients used in the teaching exercises were not those used in the exercise itself.)

The order of assessment was randomized according to a 7x7 Latin square design.

In this exercise the MITAX assessed muscle disease, mucocutaneous disease, skeletal disease, gastrointestinal disease and cardiovascular/respiratory disease. Visual analogue scales (MYOACT) were used to assess the same organs and/or systems, with the exception that the cardiovascular and pulmonary systems were assessed separately. In addition the physicians were

asked to score 'any other disease activity' using the visual analogue scale.

The MDI was used to assess damage in the ocular, pulmonary, cardiovascular, peripheral vascular, gastrointestinal, musculoskeletal, cutaneous and endocrine organs and/or systems. The physicians were also asked to score 'any other' damage they found. A total damage score was calculated by summing the individual scores. Visual analogue scales (MYODAM) were used to assess damage in the same systems, with the exception that the muscular and skeletal organs and/or systems were assessed separately. The physicians were also asked to give a global score using the visual analogue scale.

Following this exercise detailed discussions took place. Dissatisfaction with some aspects were expressed, in particular, it was felt that it was necessary to add a constitutional element to the MITAX to capture some aspects of disease activity that, although uncommon, were felt to be important by many. The majority of the participating physicians were also unhappy using a combined pulmonary and cardiovascular measurement in the MITAX. Further discussions then took place by e-mail and the following changes were agreed.

A constitutional element was added to the MITAX and the MITAX cardiovascular/respiratory disease category was split into two separate categories, the cardiac and the pulmonary. This re-categorization also included a certain amount of revision of the areas physicians were asked to consider. In particular a number of areas were split into smaller subcategories. For example, dyspnoea was initially considered only at rest; subsequently physicians were asked to consider it at rest and on exertion. In addition the following minor changes were made to the MITAX. In the muscle disease category, the first version simply required physicians to classify loss of function into major or moderate. In the second version they were asked to classify loss of function

44

into major, moderate or mild. Similarly in the skeletal disease category, in the first version, physicians were asked to classify arthritis into either severe polyarthritis with major loss of function or arthritis with moderate loss of function. In the second version a third category was added allowing arthritis also to be classified as mild/no loss of function. In the mucocutaneous category, the first version of the MITAX asked physicians to consider 'vascular changes resulting in necrosis'; this is omitted from the second version. In the gastrointestinal category, the first version required the physician to consider mild/moderate abdominal pain. In the second version the physician is asked to consider severe abdominal pain, moderate abdominal pain and mild abdominal pain separately.

In addition the MDI musculoskeletal category was split into a muscular and a skeletal category. No other changes were made.

Subsequently in May 2002, a second real patient exercise took place. The order of assessment was again randomized according to a 7x7 latin square, and again the consultations lasted for up to one hour. Prior to attending the second exercise the physician-assessors reviewed training materials, as well as scoring MITAX, MYOACT, MDI and MYODAM and viewing a cutaneous slide collection covering example ratings for each instrument and manual muscle strength testing procedure.

For pragmatic reasons three of the patients and six of the physicians were the same for both exercises. The three patients that took part in both exercises all had active disease, and, to provide a variety in the symptoms being assessed, were chosen for the second exercise as well as the first. It was unfortunate that it was not possible to find a different set of physicians for the second exercise, but the research is only of interest to a limited number of physicians as the disease is rare, and it is also difficult to get attendance

45

for a whole weekend.

The implications of this, for the analysis of the real patient exercises and the interpretation of the results, are relatively minor, with respect to the characterization of the use of the instruments across physicians. The assessments took place about 10 months apart, rendering it most unlikely that the physicians would recall the patient's previous activity/damage in any detail.

The need for precise and accurate measurement

Accurate or precise measurement is an important component of any study design, and there is a need here to assess the extent to which the measurements are subject to error and the degree to which clinical scores might meaningfully represent patient status.

In medical research it is of interest to examine the extent to which the results of a classification procedure concur in successive applications. In this situation, as with many in medical practice and research, the measurements are based on observations made by clinicians and are clearly prone to individual variation in observation practice. Here, as in any measurement situation, besides intra-observer and inter-observer disagreements variability among patients may arise for many reasons.

A generalizability study (Cronbach *et al.* 1972) collects data from which estimates can be made of the components of variance for measurements made by a certain procedure. In a generalizability study one obtains two or more scores for a person, and examines the consistency of the scores.

An important goal of generalizability studies is to identify and measure variance components that contribute errors to a measurement. They provide information on observer reliability and agreement, identifying sources of error

46

and determining the relative importance of each component. This preliminary testing of the myositis tools can be viewed as a generalizability study, and should be regarded as an integral part of the instrument development.

The design of the real patient exercises, and an approach specifically developed for the study of the reliability of the tools and the level of agreement amongst the physicians, are described in sections 3.2 and 3.3. The results from both real patient exercises are presented with an overview in section 3.4.

The main purpose of the statistical analysis of the experiments is to assess the reliability of the tools and the agreement among the physicians. Since it is not possible to have a valid measure unless the measure has some degree of reliability, demonstration of reliability is generally viewed as a necessary first step to establishing the quality of a measure (Shrout (1998), Carey and Gottesman (1978)). It is to be hoped that the results of the statistical analyses are of use in indicating those areas that require further discussion. The analysis of the first exercise was an interim analysis. Both exercises provide information on the behaviour of the tools.

3.2.1 Experimental Design

The Latin square design eliminates two extraneous sources of variation by using two-way or double blocking on the experimental units. A JxJ Latin square, is a square matrix with J rows and J columns. Each of the resulting J^2 cells contains one of J letters.

In this experiment each letter corresponds to one of the patients and each letter occurs once and only once in each row and column. Each column represents one of the physicians and each row represents the time at which the patient was seen.

The model used is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \tau_k + \epsilon_{ijk} \qquad (3.1)$$

for $i, j, k = 1, 2, ..., J$. Here y_{ijk} denotes the ith rating ($i = 1...J$) on the the jth patient ($j = 1....J$) at the kth time; $-\infty < \mu < \infty$ is the overall population mean of the ratings, α_i is the difference from μ of the mean of the ith physician's rating (ie the mean across measurements made on all targets), β_j is the difference from μ of the jth patients so-called true score (i. e the mean across many repeated ratings on the jth target), τ_k is the effect of the kth time slot and ϵ_{ijk} is the random error.

The model is additive, any interaction between time, physicians or patients is identified with error. Furthermore, since there is only one observation in each cell, only two of the subscripts i, j, and k are needed to denote a particular observation. Assuming that the patients are randomly selected from a larger population, equation (3.1) is appropriate in two different situations. The first is one where a random sample of J physicians are selected from a larger population, and each physician rates each patient (random effects), and the second is where each patient is rated by each of the same J physicians who are the only physicians of interest (mixed effects).

For a $J \times J$ Latin Square the analysis of variance consists of partitioning the total sum of squares of the $N = J^2$ observations into appropriate components, here, time, physicians, patients and error by using the identity

$$y_{ijk} - \bar{y}_{...} = (\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) + (\bar{y}_{..k} - \bar{y}_{...})$$
$$+ (y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{..k} + 2\bar{y}_{...}).$$

Squaring each side and summing over i, j, k and noting that (i,j,k) take on only J^2 values leads to

$$SS_T = SS_R + SS_C + SS_\beta + SS_E,$$

48

where

$$SS_T = \sum_{i=1}^{J} \sum_{j=1}^{J} \sum_{k=1}^{J} (y_{ijk} - \bar{y}...)^2$$

$$SS_R = J \sum_{i=1}^{J} (\bar{y}_{i..} - \bar{y}...)^2$$

$$SS_C = J \sum_{j=1}^{J} (\bar{y}_{.j.} - \bar{y}...)^2$$

$$SS_\beta = J \sum_{k=1}^{J} (\bar{y}_{..k} - \bar{y}...)^2$$

and

$$SS_E = \sum_{i=1}^{J} \sum_{j=1}^{J} \sum_{k=1}^{J} (y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{..k} + 2\bar{y}...)^2.$$

(SS_T is the total sum of squares, SS_R is the sum of squares due to the rows (that is time), SS_C is the sum of squares due to the columns (that is the physicians), SS_β is the sum of squares due to the patients and SS_E is sum of squares due to error.) The corresponding degrees of freedom are partitioned as

Total		Rows		Columns		Treatments (patients)		Error
$J^2 - 1$	=	$(J-1)$	+	$(J-1)$	+	$(J-1)$	+	$(J-1)(J-2)$.

The ϵ_{ijk}'s and the time component τ_k's are normal random variables with means zero and variances σ_ϵ^2 and σ_τ^2 respectively. Assuming that the patients are randomly chosen from a larger population the patient component β_j is also a normal random variable with mean zero and variance, σ_β^2.

If the J physicians are assumed to be a random sample selected from a larger population then α_i is a random variable that is assumed to be normally

distributed with a mean zero and variance σ_α^2. If however the J physicians are assumed to be the only physicians of interest α_i is a fixed effect subject to the constraint $\sum_{i=1}^{J} \alpha_i = 0$. The expected mean square component corresponding to σ_α^2 is $\theta_\alpha^2 = \sum_{i=1}^{J} \alpha_i^2 / (J-1)$. The complete analysis of variance for both cases is shown in table 3.1.

Table 3.1: The Complete Analysis of Variance for model 3.1

Source of Variation	Degrees of Freedom	Sums of Squares	Mean Square	Expected Mean square (random effects)	Expected Mean square (mixed effects)
Physician	$J-1$	SS_{Phys}	MS_{Phys}	$\sigma_\epsilon^2 + J\sigma_\alpha^2$	$\sigma_\epsilon^2 + J\theta_\alpha^2$
Patient	$J-1$	SS_{Pat}	MS_{Pat}	$\sigma_\epsilon^2 + J\sigma_\beta^2$	$\sigma_\epsilon^2 + J\sigma_\beta^2$
Time	$J-1$	SS_{Time}	MS_{Time}	$\sigma_\epsilon^2 + J\sigma_\tau^2$	$\sigma_\epsilon^2 + J\sigma_\tau^2$
Error	$(J-1)(J-2)$	SS_E	MS_E	σ_ϵ^2	σ_ϵ^2
Total	J^2-1	SS_T			

3.3 Agreement and Observer Reliability

The terms 'observer reliability' and 'agreement' are often used interchangeably, but in theory they are different concepts.

Reliability coefficients express the ability to differentiate among subjects. They are ratios of variances: in general, the variance attributed to the difference among subjects divided by the total variance. If for example a test were applied at two time points and the data from successive applications were available, ideally the results would be the same. However variation in the method and location of sampling as well as variation in other (laboratory) procedures may lead to different outcomes. In this context it might be said that there is empirical evidence that the test is reliable if the majority of the subjects are classified in the same way for both applications of the test. Empirical evidence of an unreliable test may lead to refinements of the testing procedure.

Agreement refers to conformity. Agreement parameters determine whether the same value is achieved if a measurement is performed more than once, either by the same observer or by different observers. Observers would typically be said to exhibit a high degree of agreement if a high percentage of their measurements concurred, and poor agreement if they often made different measurements. In general the latter outcome could arise if the categories were ill-defined, if the criteria for assessment were different for the observers or if the ability of the observers to examine the criteria differed sufficiently, possibly as a result of different training or experience. Poor empirical agreement might therefore lead to a review of the category definitions and diagnostic criteria, or possibly retraining with a view to improving agreement and hence consistency of diagnoses and treatment.

In a more heterogeneous population (with larger ranges) the value of a

reliability coefficient will be higher reflecting the fact that in heterogeneous populations subjects are easier to distinguish than in homogeneous populations. It might be imagined that in a heterogeneous population reliability and agreement measures will correspond well. In homogeneous populations however generally reliability is low as it is difficult to distinguish between patients, but agreement may well be high (Stratford 1989). Caution should be exercised in comparing agreement parameters between populations and in extrapolating results on these parameters to populations that differ in respect to heterogeneity (Stratford 1989).

For the experiment in question it is of interest not only to measure how well the newly developed tools differentiate between patients but also to attempt to assess the level of agreement between physicians rating the same patient using these tools.

A large number of outcomes are to be examined, and although a number of measures could be used to assess agreement and reliability it was decided that a commonality in analysis would be useful. The MYOACT and MYODAM scores are on a continuous scale from 0-100, but the manual muscle scores, the MITAX scores and the total MDI are all discrete. However the integer responses could be treated as if they came from a Gaussian general linear model, and as a result of this, as a first approach to the assessment of agreement and reliability of the Myositis tools, an Intra Class Correlation Coefficient was used.

3.3.1 The Intraclass Correlation Coefficient

The Intra Class Correlation Coefficient (ICC) is primarily a measure of reliability designed for continuous variables, although ordinal data is sometimes treated as continuous for the purpose of the calculation of an ICC. It is de-

fined as the correlation between one measurement (either a single rating or a mean of several ratings) on a target and another measurement obtained on that target. The intraclass correlation, originally developed by Pearson, was made a part of the analysis of the variance components by Fisher. He states in 'Statistical Methods for Research Workers' (Fisher 1925) that a very great simplification is introduced into questions involving intraclass correlation when it is recognized that in such cases the correlation merely measures the relative importance of two groups of factors causing variation.

The ICC is a bona fide correlation coefficient that is often but not necessarily identical to the component of variance due to targets divided by the sum of it and other variance components. If the group means are assumed to be distributed with a variance component σ_b^2 and the within-group deviations with a component σ_w^2 then the usual definition of ρ_I the population value for the ICC is given by

$$\rho_I = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2},$$

(3.2)

where ρ_I ranges from 0 to +1.

The ICC is, like any other correlation coefficient, dependent on the range of the variables measured. With larger ranges, that is, a more heterogeneous population the value of the ICC is higher. This reflects the fact that, as previously mentioned, in heterogeneous populations subjects are easier to distinguish than in homogenous populations.

However although the ICC is designed to measure how well patients can be distinguished from each other despite measurement errors, it is also used as a measure of agreement. It can be seen from equation (3.2) that a high value for the ICC is naturally associated with a small within subjects variance. For example if a set of judges agree perfectly about a set of subjects, making for a within subjects variance of zero (perfect agreement) ICC $= 1$ which also

indicates total reliability.

The results of an analysis of variance of experimental data can, if required, be used (and will in this experiment) to estimate the variance components attributable to different classes of effect. A description of the estimation of the components of variance from an ANOVA in its general form is as follows. Let $\boldsymbol{\sigma}^2$ be the vector of variance components to be estimated in some model and let $\boldsymbol{s}$ be a vector of sums of squares. Then, when each sum of squares has an expected value that is a linear function of the variance components, $E(\boldsymbol{s})$ is a vector of such linear functions that we can represent as $\boldsymbol{C}\boldsymbol{\sigma}^2$, so that

$$E(\boldsymbol{s}) = \boldsymbol{C}\boldsymbol{\sigma}^2. \tag{3.3}$$

Hence for non-singular $\boldsymbol{C}$ the ANOVA estimator of $\boldsymbol{\sigma}^2$ is based on (3.3) and is the solution for $\hat{\boldsymbol{\sigma}}^2$ to

$$\boldsymbol{s} = \boldsymbol{C}\hat{\boldsymbol{\sigma}}^2, \tag{3.4}$$

namely

$$\hat{\boldsymbol{\sigma}}^2 = \boldsymbol{C}^{-1}\boldsymbol{s}. \tag{3.5}$$

It is clear from (3.5) that each element of $\hat{\boldsymbol{\sigma}}^2$, i.e. each estimated variance component, is a linear component of the sums of squares in $\boldsymbol{s}$. It is important to note however that there is nothing inherent in (3.5) to ensure that every element of $\hat{\boldsymbol{\sigma}}^2$ is always non-zero. Thus it is that ANOVA estimates for parameters that are by definition positive can be negative.

Generally it is accepted that a negative estimate either may be indicative of a wrong model or an indication that the true value of the variance component is zero (Searle *et al.* 1992). However Nelder (1954) states that there is no justification for the common practice of putting equal to zero

a variance component whose estimate from an analysis is negative. Nelder states that the appropriate formulae should be applied as though the component were negative and special distinction should be made. In line with this Fleiss (1985) states that it would be a mistake to "correct" the estimate by changing it to zero, because the effect on the estimation of ρ_I would be systematically to bias the intraclass correlation coefficient. Sitgreaves (1960) points out that it is also possible, even with positive variance estimates, when each target is rated by each of the same k judges who are the only judges of interest, for the population value of the ICC to be negative. (A negative ICC is usually taken to be zero reliability).

There are numerous versions of the ICC that can give quite different results when applied to the same data. The choice of the appropriate form of the ICC depends on the choice of the appropriate statistical model and to the potential use of its results. Shrout and Fleiss (1979) state that the choice of the appropriate form of the ICC calls for three decisions; a) is a one-way analysis of variance appropriate for the analysis of the reliability study or should judge leniency or severity be treated as a source of error? b) Are differences between the judges' mean rating relevant to the reliability of interest. c) Is the unit of analysis an individual rating or the mean of several ratings? The first and second pertain to the appropriate statistical model for the study, and the second and third to the potential use of its results.

Given that in this experiment each physician rates each patient, a one way analysis of variance would not have been appropriate, and, given our desire to assess the agreement between the physicians, it is certainly of interest to partition the within-target sum of squares into a between-physicians sum of squares and a residual sum of squares.

3.3.2 An Intraclass Correlation Coefficient for the Myositis Real Patient Exercises

The model used in both real patient exercises is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \tau_k + \epsilon_{ijk};$$

the covariance between two ratings on a patient is σ_β^2 and the total variance is $\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\tau^2 + \sigma_\epsilon^2$. Given the fact that the usual definition of the population value for the ICC is

$$\rho_I = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2}, \qquad (3.6)$$

a possible formula for estimating ρ_I for this experiment might therefore be

$$\rho_I = \frac{\sigma_\beta^2}{\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\tau^2 + \sigma_\epsilon^2}. \qquad (3.7)$$

However there are two problems with this definition of ρ_I. Firstly it is not clear that equation (3.7) is in fact a correlation coefficient and, secondly, although time has been adjusted for in the analysis it can clearly be regarded as an artefact of the design of the experiment and it would be preferable to assume that all the patients were seen at the same time. If this assumption were made, then no unbiased estimator of σ_β^2 would be available if there was any interaction between time and patients. However if there was some interaction between time and patients then σ_β^2 would no longer be the covariance between ratings on a target because of correlated interaction terms, and the actual covariance would be equal to $\sigma_\beta^2 - \frac{\sigma_{Int}^2}{J-1}$ where σ_{Int}^2 is the variance of the interaction. The total variance would then be $\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\epsilon^2$ giving ρ_I the population value for the ICC as

$$\rho_I = \frac{\sigma_\beta^2 - \frac{\sigma_{Int}^2}{J-1}}{\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\epsilon^2}.$$

56

However in using a Latin Square for the design of the experiment it was assumed that there was no interaction between patients, physicians and time and therefore in this case $\sigma_{Int}^2 = 0$ and σ_β^2 is the covariance between the ratings on a patient. This gives, for this experiment, ρ_I the population value for the ICC as

$$\rho_I = \frac{\sigma_\beta^2}{\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\epsilon^2}.$$

This can then be estimated consistently but with bias by estimating the variances from the ANOVA as follows;

$$\sigma_\beta^2 = \frac{MS_{Pat} - MS_E}{J}$$

$$\sigma_\alpha^2 = \frac{MS_{Phys} - MS_E}{J}$$

$$\sigma_\epsilon^2 = MS_E.$$

Consequently when the physicians are assumed to be selected at random from a larger population ρ_I is estimated by

$$ICC(1) = \frac{MS_{Pat} - MS_E}{MS_{Pat} + MS_{Phys} + (J-2)MS_E}. \tag{3.8}$$

In the case where the physicians are assumed to be the only physicians of interest ρ_I is estimated (consistently but with bias) by

$$ICC(2) = \frac{MS_{Pat} - MS_E}{MS_{Pat} + (J-1)MS_E}. \tag{3.9}$$

As in this instance we would hope that the new tools can be effectively used by many physicians, and consequently wish to generalize the results to other physicians within some population, the physicians should be considered to be a sample from a larger population and we should use ICC(1).

It should be noted that both ICC(1) and ICC(2) give the expected reliability of a single judges rating, which is appropriate in the experiment. If the unit of reliability was the mean of a number of ratings, the ICC's would need to be altered to allow for this (Shrout and Fleiss 1979).

3.3.3 The Construction of a Confidence Interval for ICC(1)

The hypothesis that $\rho_I = 0$ is equivalent to the hypothesis that $\sigma_\beta^2 = 0$. This hypothesis is rejected if the ratio $\frac{MS_{pat}}{MS_E}$ exceeds the critical value of the F distribution with (in this experiment) $(J\text{-}1)$ and $(J\text{-}1)(J\text{-}2)$ degrees of freedom. A confidence interval for ρ_I however must be a function not only of MS_{pat} and MS_E but of MS_{phys} as well. If ρ_I were known but the individual components of variance not, following Fleiss and Shrout (1978) the expectation of MS_{pat} could be expressed as

$$E(MS_{pat}) = \frac{1}{1 - \rho_I}(J\rho_I\sigma_\alpha^2 + [1 + (J - 1)\rho_I]\sigma_\epsilon^2) \qquad (3.10)$$

with J being the number of physicians (and in this experiment patients) in the study. Consider

$$V = \frac{1}{(1 - \rho_I)}(\rho_I MS_{phys} + [(1 + (J - 1)\rho_I) - \rho_I]MS_E)$$

that is distributed independently of MS_{pat}. The expectation of V can be expressed as

$$E(V) = \frac{1}{1 - \rho_I}(J\rho_I\sigma_\alpha^2 + [1 + (J - 1)\rho_I]\sigma_\epsilon^2)$$

that is, equal to $E(MS_{pat})$. Satterthwaite (1946) gives the following formula for estimating the degrees of freedom of $MS = \sum_{i=1}^{k} a_i MS_i$. If MS is

approximately distributed as $\sigma^2 \chi_f^2 / f$, then

$$\hat{f} = \frac{(MS)^2}{\sum_{i=1}^{k} (a_i M S_i)^2 / f_i}. \tag{3.11}$$

Following Satterthwaite (1946), with $F_{phys} = \frac{MS_{phys}}{MS_E}$, V can be shown to be approximately distributed as $\frac{c\chi_\nu^2}{\nu}$ where χ_ν^2 denotes a variable distributed as chi square with ν degrees of freedom, $c = E(MS_{pat})$, and

$$\nu = \frac{(J - 1)(J - 2)(\rho_I F_{phys} + 1 + \rho_I (J - 2))^2}{(J - 1)J\rho_I^2 F_{phys}^2 + (1 + \rho_I (J - 2))^2}. \tag{3.12}$$

Thus the random variable $\frac{MS_{pat}}{V}$ has approximately an F distribution with $(J\text{-}1)$ and ν degrees of freedom. Let ν be estimated from (3.12) with $\hat{\rho}_I$ defined in (3.8), replacing ρ_I, then the approximate probability statement

$$1 - \alpha \approx pr(\frac{MS_{pat}}{V} < F^*),$$

where F^* is the upper $100(1 - \alpha)$ percentile of the F distribution with $(J\text{-}1)$ and ν degrees of freedom, may be converted into the approximate $100(1 - \alpha)$ percent confidence interval,

$$\rho_I > \frac{(MS_{pat} - F^* MS_E)}{F^*(MS_{phys} + (J - 2)MS_E) + F_* MS_{pat}} = \rho_{IL}. \tag{3.13}$$

An approximate confidence interval bounded above is of the form

$$\rho_I < \frac{(F_* MS_{pat} - MS_E)}{MS_{phys} + (J - 2)MS_E + F_* MS_{pat}} = \rho_{IU}, \tag{3.14}$$

where F_* is the upper $100(1 - \alpha)$ percentile of the F distribution with ν and $(J\text{-}1)$ degrees of freedom. Approximate two-sided intervals may be derived from equations (3.13) and (3.14) by using the upper $100(1 - \frac{\alpha}{2})$ percentiles.

The accuracy of Satterthwaite's approximation to the distribution of a linear combination of independent mean squares has been studied by Gaylor and Hopper (1969) and Fleiss (1971) and when the coefficients of the mean squares are all positive the approximation has been found to be good.

59

3.3.4 Disadvantages of the Intraclass Correlation Co-efficient

There are two major difficulties with the interpretation of an ICC. Firstly it is a ratio of variances and therefore difficult to interpret clinically. Secondly although sometimes used as a measure of agreement there are difficulties associated with it's interpretation as such. Once the ICC has been calculated it is not evident how much of the within subjects variance is attributable to σ_α^2 (that is to the physicians themselves). Although it is clear that a high value of an ICC is naturally associated with a small within subjects variance $(\sigma_\alpha^2 + \sigma_\epsilon^2)$, and therefore indicates good agreement, and that a small within subjects variance should yield a high intraclass correlation coefficient, indicating a high degree of reliability, it is true as previously stated, that in a homogeneous population reliability might be low while agreement is high. It is not possible to ascertain from the value alone whether a low ICC indicates low reliability and poor agreement or whether agreement might in fact be high with most of the within-subject variance due to random error. It is also possible to obtain a relatively high ICC indicating reasonably high reliability with most of the within-patient variance being due to the physicians not random error, indicating poor agreement.

In this experiment, with a particular area of interest being the amount of the within subject variance attributable to the physicians themselves, it is clear that ICC(1) alone does not provide adequate information on agreement amongst the physicians.

A possible measure of interest might therefore be the ratio of the standard error of measurement attributable to the physicians and the standard error

of measurement attributable to the patients themselves. That is

$$r = \frac{\sigma_\alpha}{\sigma_\beta}. \tag{3.15}$$

Here a small value is associated both with a small value of σ_α that is, a high level of agreement between the physicians and a high value of σ_β. This measure would enable us to comment on the level of physician agreement irrespective of the amount of variation due to random error. It is clear that r is primarily a measure of agreement, and thus the combination of r and ICC(1), would enable us to assess not only the ability of the tools to differentiate between subjects, but how much the physicians agreed.

In the case where the physicians are assumed to be selected at random from a larger population r can be estimated by

$$\hat{r} = \sqrt{\frac{MS_{phys} - MS_E}{MS_{pat} - MS_E}} \tag{3.16}$$

3.3.5 A Confidence Interval for $r = \frac{\sigma_\alpha}{\sigma_\beta}$

The hypothesis that $r = 0$ is equivalent to the hypothesis that $\sigma_\alpha^2 = 0$. This hypothesis is rejected if $\frac{MS_{phys}}{MS_E}$ exceeds the critical value of the F distribution with $(J\text{-}1)$ and $(J\text{-}1)(J\text{-}2)$ degrees of freedom. However $\hat{r}$ itself, is not a ratio of mean squares and therefore none of the standard ANOVA distributional results can be applied. As a result of this the subsequent development of a confidence interval for r follows the methodology used by Fleiss and Shrout (1978) as presented in section 3.3.3.

A confidence interval for $\frac{\sigma_\alpha}{\sigma_\beta}$ needs to be a function of not only MS_{phys} and MS_E but of MS_{pat} as well.

If r were known, but the individual components of variance were not, the expectation of MS_{phys} could be expressed as

$$E(MS_{phys}) = Jr^2\sigma_\beta^2 + \sigma_\epsilon^2, \tag{3.17}$$

61

where J is, as previously, the number of patients and the number of physicians in the study. Consider

$$V^* = r^2 MS_{pat} + (1 - r^2) MS_E \qquad (3.18)$$

that is distributed independently of MS_{phys}. The expectation of V^* can also be expressed as

$$E(V^*) = Jr^2\sigma_\beta^2 + \sigma_\epsilon^2$$

that is, equal to $E(MS_{phys})$. V^* is distributed independently of MS_{phys} but not exactly as a constant times a chi-square variable. However following Satterthwaite (1946) V^* can be shown to be approximately distributed as

$$V^* \approx \frac{\chi_f^2 E(V^*)}{f} \qquad (3.19)$$

where

$$f = \frac{(J-1)(J-2)(r^2 MS_{pat} + (1 - r^2)MS_E)^2}{(J-2)r^4 MS_{pat}^2 + (1 - r^2)MS_E^2}. \qquad (3.20)$$

Thus the random variable $\frac{MS_{phys}}{V^*}$ has approximately an F distribution with (J-1) and f degrees of freedom. Let f be estimated from equation (3.20) with $\hat{r}^2$, defined in equation (3.16), replacing r^2. Then the approximate probability statement

$$1 - \alpha \approx pr(\frac{MS_{phys}}{V^*} < F^*)$$

where F^* is the upper $100(1 - \alpha)$ percentile of the F distribution with (J-1) and f degrees of freedom, may be converted into the approximate $100(1 - \alpha)$ percent confidence interval

$$r^2 > \frac{MS_{phys} - F^* MS_E}{F^*(MS_{pat} - MS_E)} = r_l^2 \qquad (3.21)$$

An approximate confidence interval bounded above is of the form

$$r^2 < \frac{F_* MS_{phys} - MS_E}{MS_{pat} - MS_E} = r_u^2,$$ (3.22)

where F_* is the upper $100(1 - \alpha)$ percentile of the F distribution with f and (J-1) degrees of freedom. It follows that

$$r > \sqrt{\frac{MS_{phys} - F^* MS_E}{F^*(MS_{pat} - MS_E)}} = r_l,$$ (3.23)

and

$$r < \sqrt{\frac{F_* MS_{phys} - MS_E}{MS_{pat} - MS_E}} = r_u,$$ (3.24)

Approximate two-sided intervals may be derived from equations (3.23) and (3.24) by using the upper $100(1 - \frac{\alpha}{2})$ percentiles.

3.4 Results

The residuals from the analysis of each tool were examined and a selection of the results are presented in appendix B. Any apparent departures from normality did not seem sufficient to impact on the qualitative conclusions drawn.

Although generally, following the approach of Nelder (1954) and Fleiss (1985), in the calculations negative estimates of variance components have not been changed to 0, when the estimate of the physicians variance is negative it has been changed to 0 in equation (3.20) in order to obtain an upper bound for the confidence interval for r. In these cases the upper bound of the confidence interval is approximate. (The value taken for $\hat{r}$ in these cases is zero). In some instances however, the estimate of the upper bound of the confidence interval for r^2 remains negative despite the adjustment, and consequently no confidence interval is given.

63

The values of $\hat{r}$ and ICC(1) are well defined for all outcomes, but it should be noted that the distributional properties are best understood for the truly measures. Although the numerical values should therefore be interpreted with some caution, they should provide qualitative guidance for the comparison of the behaviour of the different tools.

Both measures have been used to classify the results from both real patient exercises into three categories. For the purpose of this classification an ICC(1)> 0.65 has been taken as high, indicating that a tool differentiates well between patients, and physician agreement has been considered to be high if $\hat{r} < 0.40$. (These boundaries are arbitrary and have been based on an evaluation of the results of the experiments. They have been defined to ease classification of the results.)

The first category consists of those tools where both the value of ICC(1) is high and $\hat{r}$ is low indicating that the tool is differentiating well between patients with a high level of physician agreement. These results have been categorized as GOOD.

The second category consists of those tools that demonstrate a good performance in only one of the two measures. These results have been categorized as GOOD*. Among these when $\hat{r}$ is low indicating a high level of agreement among the physicians, it appears that the low value of ICC(1) is generally due to little or no variation among the patients, and these tools can be considered to be performing reasonably well. However, in this category, when ICC(1) is high, indicating an ability to differentiate between patients, the high value of $\hat{r}$ indicates that there remained some variability among the physicians.

The third category consists of those tools where both ICC(1)is low and $\hat{r}$ is high, indicating that the tool is not differentiating well between patients

64

and there is a poor level of physician agreement. These tools have been classified as POOR.

In the first exercise there is little evidence of any patients having any 'gastrointestinal', 'cardiovascular' or 'other' activity or 'cardiovascular', 'peripheral vascular' or 'other' damage. The groups of patients did not differ enough in these areas to assess how effectively the tools detected any differences, or to draw any generalizable conclusions on the level of physician agreement. The results for these areas will not be presented.

In the first real patient exercise following a training session (using a separate patient) the manual muscle tests exhibited in general reasonable reliability and reasonable levels of physician agreement. The main areas of concern were the neck and wrist elements where, in particular, physician agreement seemed poor. However it was felt that the results for the manual muscle tests were generally satisfactory and the assessments were not repeated in the second exercise.

In the second exercise there is little evidence of any 'gastrointestinal' or 'cardiovascular' activity, or 'peripheral' vascular or 'other' damage. The groups of patients did not differ enough in these areas to assess how effectively the tools detected any differences, or to draw any generalizable conclusions on the level of physician agreement.

One patient failed to turn up for the second real patient exercise introducing some correlation between order and physician. However given that in both the first and second exercise variation due to order was very small it was felt that this correlation had a minimal impact on the results.

An overview of the findings is presented in section 3.4.1 with illustrative observations from both experiments found in appendix C. All computing was done in S-plus. Details of the routines used are given in appendix A.

Table 3.2: Estimates and 95% confidence intervals ([,]) based on results from the first real patient exercise.

	GOOD		GOOD*		POOR	
	ICC	$\hat{r}$	ICC	$\hat{r}$	ICC	$\hat{r}$
MANUAL MUSCLE						
Deltoid	0.72[0.45,0.93]	0.22[0,0.67]				
Biceps	0.82[0.61,0.96]	0.12[0,0.43]				
Quadriceps	0.85[0.67,0.97]	0.04[0,0.33]				
Gluteus Medius	0.72[0.44,0.93]	0.38[0.07,0.98]				
Gluteus Maximus			0.56[0.27,0.87]	0.28[0,0.91]		
Ankle			0.28[0.04,0.72]	0		
Neck					0.062[0.31,0.90]	0.57[0.18,1.43]
MITAX						
CV Respiratory	0.92[0.80,0.98]	0.07[0,0.27]				
Mucocutaneous			0.57[0.27,0.88]	0		
Skeletal			0.52[0.24,0.86]	0.38[0,1.12]		
Gastrointestinal			0.45[0.16,0.83]	0		
Total			0.65[0.37,0.91]	0.41[0,1.07]		
Muscle					0.48[0.20,0.84]	0.62[0.05,1.62]
MYOACT						
Pulmonary	0.73[0.46,0.93]	0.22[0,0.66]				
Mucocutaneous			0.38[0.11,0.78]	0.15[0,1.03]		
Global					0.42[0.16,0.80]	0.70[0.03,1.83]
Skeletal					0.21[0.04,0.63]	1.16[0,3.11]
MDI						
Total					0.49[0.20,0.84]	0.77[0.24,1.92]
MYODAM						
Mucocutaneous	0.84[0.65,0.97]	0.14[0,0.44]				
Pulmonary	0.74[0.48,0.94]	0.28[0,0.77]				
Ocular	0.90[0.77,0.98]	0.06[0,0.28]				
Endocrine			0.59[0.30,0.89]	0.16[0,0.72]		
Gastrointestinal					0.31[0.09,0.73]	0.87[0,2.30]
Skeletal					0.15[0,0.59]	0.67[0,2.36]
Global					0.44[0.16,0.18]	0.86[0.26,2.15]
Muscle					0.57[0.27,0.88]	0.60[0.17,1.52]
OTHER						
Physician's Global Activity Score					0.40[0.14,0.79]	0.48[0,1.41]
Physician's Global Damage Score					0.51[0.21,0.85]	0.70[0.19,1.76]

Table 3.3: Estimates and 95% confidence intervals ([,]) based on results from the second real patient exercise.

	GOOD ICC	GOOD $\hat{r}$	GOOD* ICC	GOOD* $\hat{r}$	POOR ICC	POOR $\hat{r}$
MITAX						
Pulmonary			0.33[0.07,0.79]	0.22[0,1.10]		
Mucocutaneous			0.23[0,0.73]	0		
Skeletal	0.74[0.46,0.95]	0.20[0,0.61]				
Gastrointestinal			0.35[0.08,0.80]	0.15[0,1.02]		
Total					0.34[0.09,0.79]	0.62[0,1.74]
Muscle			0.33[0.06,0.80]	0		
Constitutional					0.43[0.15,0.84]	0.58[0,1.57]
MYOACT						
Pulmonary			0.45[0.16,0.85]	0.40[0,1.18]		
Mucocutaneous			0.32[0.04,0.79]	0		
Skeletal	0.65[0.35,0.93]	0.23[0,0.72]				
Muscle	0.74[0.46,0.95]	0				
Constitutional	0.71[0.41,0.94]	0.15[0,0.56]				
Gastrointestinal			0.38[0.10,0.82]	0.17[0,0.96]		
MDI						
Mucocutaneous			0.60[0.29,0.91]	0.13[0,0.63]		
Pulmonary					0.41[0.13,0.83]	0.49[0,1.41]
Ocular	0.78[0.52,0.96]	0.12[0,0.45]				
Endocrine			0.54[0.23,0.89]	0.17[0,0.74]		
Gastrointestinal			0.55[0.24,0.89]	0.34[1,1.00]		
Skeletal					0.36[0.09,0.80]	0.45[0,1.38]
Global	0.70[0.41,0.94]	0.20[0,0.64]				
Muscle	0.76[0.49,0.95]	0				
Cardiovascular	0.86[0.65,0.97]	0				
Infection			0.25[0.02,0.74]	0.24[0,1.32]		
MYODAM						
Mucocutaneous	0.70[0,0.94]	0				
Pulmonary	0.78[0.51,0.96]	0.10[0,0.43]				
Ocular	0.75[0.46,0.95]	0				
Endocrine			0.50[0.18,0.87]	0		
Gastrointestinal			0.52[0.21,0.88]	0.28[0,0.91]		
Skeletal					0.28[0.04,0.76]	0.43[0,1.48]
Global	0.67[0.37,0.93]	0.28[0,0.80]				
Muscle			0.64[0.33,0.92]	0.35[0,0.96]		
Cardiovascular	0.82[0.59,0.97]	0				
Infection			0.22[0.00,0.72]	0.23[0.1.40]		
OTHER						
Physician's Global Activity Score	0.68[0.38,0.93]	0.29[0,0.80]				
Physician's Extra Muscular Score			0.54[0.23,0.89]	0.29[0,0.91]		

3.4.1 An Overview of The Findings

Assessment of Disease Activity

In the first exercise, with respect to assessment of disease activity, the MITAX seemed to work well with, perhaps paradoxically, only the assessment of disease activity in the muscle system exhibiting poor reliability and a low level of agreement among the physicians. The MYOACT assessments exhibited both poor reliability and a low level of agreement among the physicians in the global and skeletal assessments.

In the second real patient exercise the MITAX appeared in the main to lead to a high level of agreement between the physicians. However both agreement and reliability were poor with respect to the constitutional element. The MYOACT assessments all appeared to exhibit a high level of agreement among the physicians.

In summary, the analysis of the results from both exercises suggest that both the MITAX and MYOACT perform well, with lower intraclass correlation coefficients generally being associated with a lack of variation in the patients' disease. Only the constitutional element of the MITAX (only assessed in the second exercise), the TOTAL MITAX (average ICC = 0.50, average $\hat{r}$ = 0.70) and the skeletal element of the MYOACT (average ICC = 0.43, average $\hat{r}$ = 0.52) performed poorly on average.

Assessment of Disease Damage

In the first real patient exercise the total MDI performed poorly with respect to both reliability and agreement. The MYODAM performed poorly with respect to reliability and agreement in the gastrointestinal, skeletal, muscle and global elements.

68

In the second real patient exercise there was generally a high level of agreement among the physicians for the different components of the MDI. Only the skeletal and pulmonary MDI elements performed poorly with respect to both reliability and agreement. Skeletal assessment in the MYODAM also performed poorly with respect to both reliability and agreement.

In summary the results from both exercises suggest that the MDI performs well. As the individual elements of the MDI were only assessed in the second exercise, the assessment of agreement and reliability can only be based on the results from that exercise. As previously mentioned only the skeletal and pulmonary elements perform poorly. However the MYODAM generally appears to perform less well than the MDI. The gastrointestinal (average ICC = 0.42, average $\hat{r}$ = 0.58), skeletal (average ICC = 0.22, average $\hat{r}$ = 0.55), global (average ICC = 0.55, average $\hat{r}$ = 0.57) and muscle (average ICC = 0.61, average $\hat{r}$ = 0.48) elements, on average, all perform poorly with respect to both reliability and agreement.

There is no doubt that methods of assessing disease activity and damage in myositis are urgently required. Professor D. Isenberg and IMACs now propose that those interested in assessing myositis patients either as part of long term outcome studies or in a drug trial should use one or both of the activity or damage measure, albeit that formal validation studies in larger patient populations are being planned.

Chapter 4

Analysis of the Interrelationships between Disease Activity in the Different Organs and/or Systems in Systemic Lupus Erythematosus.

The definition of the internationally recognized disease activity and damage measures for the assessment of lupus (as described in section 2.1.3) have brought considerable order to what had become an increasingly chaotic situation, and have provided a consistent way to assess the disease. However the clinical manifestations of lupus prove a great challenge to rheumatologists managing the condition and it would clearly be of use to clinicians to be able to identify those patients at risk of developing active disease. There is a need

to better understand predictors of disease activity in order to improve and standardize therapy and to prevent the development of chronic damage.

Observation of patients with lupus suggests, to some physicians (personal communication with Professor D. Isenberg and Dr C. Gordon), that subsets exist within the disease, and it has been suggested that the different subsets may benefit from different approaches to treatment. It is of interest to attempt to assess the validity of this clinical impression.

The aim and purpose of the subsequent analysis is to help develop models for disease activity in lupus based on the BILAG system. As well as attempting to assess the validity of the hypothesized existence of subsets of the disease, a particular focus of the analysis will be on whether there exists a reliable method of identifying those patients with lupus who are most at risk of experiencing an increase in disease activity. Since it appears that some patients follow a more benign course it could be hypothesized that there may be identifiable prognostic indicators for disease severity and activity. For example, an increase in disease activity in one system might affect the rate of increased activity in the same system and may also alter the rate of activity in other systems.

The development of these models will help with the understanding of the apparent correlation between activity in different organs and/or systems and will examine the prognostic importance of activity levels. Approaches to data analysis will be developed so that these questions can be answered.

In general, for causal inferences, biological plausibility is a major criterion. However, the search for relationships in medical data may or may not have substantial biological motivation. For example, in a routine analysis of data from a marrow transplant programme, Storb et al. (1977) established a relationship between graft rejection and the amount of donor marrow used

71

in the transplant procedure. While this finding was unexpected, it neverthe-
less led to a proposal for the modification of the existing clinical procedures,
suggesting that when the number of available marrow cells fell below a cer-
tain threshold, the use of buffy-coat cells from the marrow donor might be a
reasonable approach.

The database under consideration comprises information on patients from
lupus clinics at two hospitals in Birmingham, The Queen Elizabeth Hospital
and the City Hospital. It incorporates considerable demographic information,
the BILAG activity index, the SLICC damage index and detailed information
on therapy and laboratory data. The data has been collected on 440 patients,
29 men and 411 women over a period of ten years. 277 of the patients are
Caucasian, 72 Afro-Carribean, 67 Asian and 24 of other ethnic origins.

4.1 An Introduction to the Analysis of Repeated Categorical Outcomes

Longitudinal data sets, comprised of an outcome variable, y_{it}, and a px1
vector of covariates x_{it}, observed at times $t = 1, ..., n_i$ for subject $i = 1, ..., K$,
occur frequently. Typically the interest is in either the pattern of change over
time or more simply the dependence of the outcome on the covariates.

With a single outcome for each subject ($n_i = 1$), a generalized linear
model (McCullagh and Nelder 1989) can be applied to obtain such a descrip-
tion for a variety of continuous or discrete outcome variables. With repeated
observations, however, the correlation among values for a given subject must
be taken into account. Ignoring correlation where it exists can lead to incon-
sistent estimates of precision and incorrect inferences concerning regression
coefficients.

There are therefore two main objectives for statistical models of longitudinal data; firstly, for convenience to adopt the conventional regression tools that relate the response variable to the explanatory variables, and secondly, as much as possible to account for the within subject correlation. Generally the regression objective is of primary interest, as while it is essential to account for the within-subject correlation, the nature of the correlation is often of secondary interest (Zeger and Liang 1992).

In (section 4.2) two approaches to the analysis of longitudinal data are considered in the development of methods for the analysis of the lupus data. Section 4.1.1 outlines details of these approaches.

To introduce these approaches let $y_i = (y_{i1}, ..., y_{it}, ..., y_{in_i})'$ be a n_i vector of repeated responses for the ith subject (i=1,...,K), with y_{it} and x_{it} as previously defined.

4.1.1 Marginal and Transition Models

The distinction between marginal models and models for transitions is important. For many families of non-linear models the sets of marginal and transition models are disjoint except in trivial cases.

In a marginal model the target of estimation is the population averaged or cross-sectional mean response. Marginal models give representations for the marginal distributions of the response at each occasion, and the dependence of those distributions on personal characteristics and other independent variables; they do not model individual changes over time or the effects of covariates on individual change. Marginal models inform us about the average state of a population and tell us nothing about the relationships among the responses for individual members of the population. For example two responses may have identical marginal distributions without there being a

similar dependence between the responses for many or indeed any individual subjects.

In a transition model the target of estimation is the conditional mean at a fixed time given the history of responses to that point. Models for transition describe the distribution of individual changes over time and the effects of individual characteristics or risk factors on those changes. Models for transitions can be represented as probability distributions for the future state given the individual's history. In a transition model, the distribution of each variable is considered conditionally on previous outcomes in the sequence. They give representations for the transition probabilities between outcome states at successive occasions. The model represents the behaviour of changes from a previously established position.

Marginal models are sometimes called population averaged models, whereas transition models are termed subject specific.

The choice between the two types of model will depend on the questions being asked and the inferences that are required from the analysis. Questions addressed by marginal models include questions regarding changes in the overall prevalence of the disease and changes in the role of risk factors. Changes in marginal distributions are sometimes called 'net changes'; they are longitudinal in one sense in that they involve comparisons over time, however they could be studied with cross-sectional data, provided that the cohort effects were sufficiently small. Marginal models do not address questions concerning heterogeneity between subjects, nor in the longitudinal sense do they address questions concerning the possible effect of a subject's previous responses on their current response.

Questions addressed by transition models concern changes in an individual's state over time. Given this they can only be investigated with longitu-

dinal data. Unlike the marginal model the transition model has no representation in terms of cross-sectional data.

If one is interested in analysing conditional probabilities for a certain state or transitions to a state given the individual's history and the effect of covariates on these probabilities, then conditional models are needed. However if the main objective is the effect of covariates on the outcome variable then marginal models are often more appropriate.

As stated marginal and transition models differ in their target estimation; this has implications for the correct interpretation of the regression parameters. Marginal regression coefficients have the same interpretation as coefficients from a cross-sectional analysis. They have 'population averaged' interpretations because they contrast odds of disease in the populations with or without the risk factor. Because the coefficients describe the effect of covariates on the marginal expectation of the y's they have the same interpretation regardless of the number of repeated observations n_i, which may vary among subjects. If β is a regression coefficient in a marginal model with β^* a regression coefficient from an analogous transition model, then e^{β} is the ratio of population prevalences, whereas because of the adjustment for past responses, e^{β^*} is approximately the ratio of incidences from two groups.

Marginal Models

In this approach the regression and within subject correlation are modelled separately. It is assumed[1]:

1. the marginal expectation of y_{it}, $E(y_{it}) = \mu_{it}$, is related to x_{it} by

$$g(\mu_{it}) = x'_{it}\beta$$

where g is a known link function;

---

[1] As presented in the literature (for example in Diggle *et al.* (2002))

2. the marginal variance is a function of the marginal mean, that is,

$$var(y_{it}) = \nu(\mu_{it})\phi$$

where ν is a known function and ϕ is the over-dispersion parameter that accounts for the variation of y_{it} not explained by $\nu(\mu_{it})$;

3. the covariance between y_{is} and y_{it}, $s < t = 1, ..., n_i$ is a function of the marginal means and additional parameters $\boldsymbol{\alpha}$, that is,

$$cov(y_{is}, y_{it}) = c(\mu_{is}, \mu_{it}; \boldsymbol{\alpha}),$$

where c is a known function.

It should be noted that only the first two moments of the joint distribution of y_i are specified by 1-3. Likelihood inference requires additional assumptions, as suggested for binary data by Prentice and Zhao (1991) or Liang and Zeger (1992). As a consequence of this the likelihood is often intractable as it will involve many nuisance parameters.

Liang and Zeger (1986) developed a method for estimation based on an extension of generalized linear models to the longitudinal setting. They introduced a class of estimating equations that take the correlation into account. The resulting estimates of the regression parameters remain consistent, and in addition, consistent variance estimates are available under the weak assumption that the weighted average of the estimated correlation matrices converges to a fixed matrix. . Their methods are semi-parametric in that the estimating equations are derived without fully specifying the joint distribution of a subject's observations. Instead they specify the likelihood for the (univariate) marginal distributions and a 'working' covariance matrix for the observation vectors.

Let y_{it} have the exponential family density

$$f(y_{it}) = \exp([y_{it}\theta_{it} - a(\theta_{it}) + b(y_{it})]/\phi)$$

76

that corresponds to a generalized linear model (McCullagh and Nelder 1989), when θ_{it} is defined via a link function $\theta_{it} = h(\eta_{it})$, and $\eta_{it} = g(\mu_{it}) = x'_{it}\beta$. Liang and Zeger (1986) consider score like estimating equations of the form

$$\sum_{i=1}^{m} D'_i V_i^{-1} S_i \qquad (4.1)$$

where S_i is the Tx1 vector of deviations $y_{it} - E(y_{it})$, for the ith individual where $(E(y_{it}) = da(\theta_{it})/d\theta_{it})$, and V_i is the TxT 'working' covariance matrix for y_i and $D_i = dE(y_i)/d\beta$. V_i is represented as

$$V_i = A_i^{1/2} Corr(y_i) A_i^{1/2}$$

where $A_i(= diag(var(y_{it}))$ is specified by the marginal distributions. The estimates of the regression parameters are obtained by specifying a model for $Corr(y_i)$ and solving for β iteratively. The authors show that, under mild regularity conditions, $\hat{\beta}$ is asymptotically unbiased and normal for any choice of $Corr(y_i)$, with asymptotic variance depending on both assumed and true covariance pattern. In addition, as stated, a consistent variance estimate is also available under equally weak conditions. The method allows for time dependent covariates.

Transition Models

In this approach the correlation is directly incorporated into a regression model. Parameters for the dependence on x_{it} and for correlation are introduced in a common equation. Transition models assume[2]:

1'. the conditional expectation of y_{it}, $\mu_{it}^c = E(y_{it} \mid y_{it-1}, ..., y_{i1})$, depends on x_{it} and past responses as follows:

$$g(\mu_{it}^c) = x'_{it}\beta^* + \sum_{j=1}^{\eta} \alpha_j^* f_j(y_{it-1}, ..., y_{i1}), \qquad (4.2)$$

---

[2]As presented in the literature (for example in Diggle *et al.* (2002))

where $f_j, j = 1, ..., \eta$ are known functions.

2′. the conditional variance of y_{it} given the past is a function of μ_{it}^c, that is

$$var(y_{it} \mid y_{it-1}, ..., y_{i1}) = \nu(\mu_{it}^c)\phi,$$

where ν is a known function.

Due to the conditional form of the transition model, the associated likelihood is typically easy to construct using the chain rule for probabilities. As a consequence the models are often comparatively straightforward to use in practice.

4.2 Possible Approaches to The Analysis of the Lupus Data

This section presents the rationale behind the choice of modelling approach used for the analysis of the lupus data.

The primary goal of the analysis being undertaken is to develop regression models that enable questions regarding interrelationships between disease activity in the different organs and/or systems[3] and correlates of disease activity to be answered. The approach that will be taken to attempt to do this will involve the definition of dynamic covariates (Aalen *et al.* 2003), that is, covariates depending on the past of the process. Dynamic covariates are continuously updated as time goes by and should sum up important aspects of the previous development of the process that may contain prognostic information. Dynamic covariates will be used here to represent disease history

---

[3]In all subsequent text organs and/or systems will be written as organs/systems and organ and/or system and organs and/or system, as organ/system. No attempt will be made to distinguish between the two.

in each of the individual organs/systems and will be included in the models for disease activity for all of the organs/systems.

The BILAG index assigns one of five possible scores to disease activity in eight organs/systems, and can be regarded as a multivariate multi-state process where the five scores correspond to states within eight correlated univariate processes. Over time patients will move between the various states, sometimes with activity increasing, other times decreasing.

Given the nature of the data, a reasonable first attempt to the modelling of the lupus data can be based on conditional (transition) models that examine the dynamic nature of the disease process by modelling the rates at which patients make transitions among the states (as represented by the BILAG index). As for many of the stochastic processes investigated in longitudinal studies, transitions between these states occur in continuous time rather than at discrete time points.

When complete records of the transitions and their times of occurrence are available, methods for the analysis of continuous-time stochastic processes are directly applicable. However, as in many longitudinal studies, because the monitoring of the lupus patients is not continuous, information is unavailable between follow up times, and the exact times of transition from one state to another are not known. The analysis must therefore be appropriate for such panel data.

A number of approaches for the analysis of panel data have been proposed. In particular Singer and Spilerman (1976) have written extensively on the problem of estimating transition probabilities for continuous-time stationary processes from discrete time data. Singer (1981) extends these methods to include some types of non-stationary processes. One of the first unified methods for entering covariates into a continuous-time model for discrete time

data is given by Kalbfleisch and Lawless (1985) who developed an efficient algorithm for obtaining maximum likelihood estimates of the transition rates from panel data under a time homogeneous Markov assumption. Satten (1996) and Goggins et al. (1998) proposed methods for applying the Cox proportional Hazard Model for the analysis of panel data. Turnbull (1976) first described interval censored data and provided a method for estimating the empirical survival function and Finkelstein (1986) proposed the use of the full likelihood proportional hazards model for regression analysis of interval-censored data. Goggins et al. (1999) solved the problem of interval censored covariates by using a Monte Carlo EM algorithm (Wei and Tanner 1990) to multiply impute exact times for each covariate observation.

However the approach previously discussed for the analysis of the inter-relationships between disease activity in the different organs/systems involving the definition of dynamic covariates would lead to the inclusion in the models of a number of interval censored internal time dependent covariates. None of methodologies mentioned can be easily extended to allow for time-dependent covariates that are interval censored when the response itself is interval censored. So although these methodologies could be used to explore the relationships between disease activity in any of the organs/systems and non-time dependent covariates, they could only be used for an analysis of the interrelationships between the organs/systems by making a number of simplifying assumptions that might not be appropriate.

A number of preliminary analyses were undertaken, making simplifying assumptions about the actual times of the events. These analyses indicated that there are relationships between disease activity in the various organs/systems and the results motivated the further detailed study discussed here. However, given the nature of the assumptions, the numerical

results of these analyses are not informative and have consequently not been presented.

Given the difficulties mentioned with the analysis of the disease process in continuous time, that clinic visits are well defined, and the obvious interest of any doctor in the state of a patient at a clinic visit, it was decided to take a very pragmatic approach to the problem and to focus on the analysis of a patient's disease state at a clinic visit.

Although it has been decided to take this pragmatic approach to the analysis of the data, and not to consider the disease process in continuous time, further discussion of the relative efficiency of this approach in relation to the use of models of the underlying process in continuous time, will be made in chapter 6.

4.2.1 The Lupus Data and Modelling a Patient's State at a Clinic Visit

A patient's state at a clinic visit is defined only at discrete time points, and in this setting the objective can be regarded as the modelling of the transition probabilities between pairs of states. In a transition generalized linear model (section 4.1) the target of estimation is the conditional mean at a fixed time point given the history of the responses to that point. In the first order transition model $\eta = 1$ (equation 4.2) and operationally, this can be achieved by including the previous state in the probability model as an independent variable. Notationally it is now convenient to regard t simply as an index for visit number. Thus given responses y_{it} and covariates x_{it} at times t=1,...,T_i we can model the probability functions,

$$\Pr[y_{it} \mid y_{i,t-1}, x_{it}] \quad t = 2, ..., T_i.$$

It is possible to classify disease activity in lupus as an approximate two state process where disease is considered to be active if the patient has BI-LAG scores A or B, and not active if the patient has BILAG scores C,D or E. Logistic regression (generalized linear model (McCullagh and Nelder 1989)) is a modelling approach that can be used to describe the relationship of several predictor variables to a dichotomous dependent variable. Consequently simple first order transition models, with t'_{ip} being the actual time of the ith patient's pth visit.

$$\text{logit}\,\Pr\left(y_{it} = 1 \mid y_{it-1}\right) = \alpha_0 + \alpha_1 y_{it-1} + \beta x_{it} + \delta(t'_{ip} - (t'_{ip-1})) \qquad (4.3)$$

$$(\text{where logit}\,\Pr\left(y_{it} = 1 \mid y_{it-1}\right) = \log \frac{\Pr\left(y_{it} = 1 \mid y_{it-1}\right)}{1 - \Pr\left(y_{it} = 1 \mid y_{it-1}\right)})$$

were used for all initial analyses. $y_{it} = 1$ if the patient scores a BILAG A or B (active disease) at a visit and $y_{it} = 0$ if a patient scores a BILAG C,D or E (not active disease) at a visit. The vector of regression coefficients x_{it} will include among other things the patient's disease history.

Although the use of the logistic models (if they are correct), enables us to model the transition probabilities between states, it is possible that there may be additional heterogeneity in the transition rates over and above that captured by the model.

Transition models can be fit using generalized estimating equations (Ware *et al.* (1988) and Korn and Whitmore (1979)), to adjust for any dependence not captured by conditioning on the patient's previous state. In this case fitting the models using the generalized estimating equation approach of Liang and Zeger (1986) will lead to population averaged estimates of the regression parameters after conditioning on the dynamic covariates.

It is also worth noting that White (1982) shows that the method of Liang and Zeger (1986) leads to consistent estimates of the variances of first order

conditional distributions and thus the estimates can provide a useful and intuitively reasonable summary of the probability distribution of individual changes even when there is some evidence for dependency on the more distant past.

One issue that might cause problems with this approach to the analysis of the lupus data is that the patients are not all seen at the same time or at regular time intervals and it will therefore be necessary to establish how much dependence there is on the time between the visits. It is clear that to a physician the state a patient presents in at a clinic visit is relevant whenever the visit may be. Nevertheless if it transpires that the time between visits has a major impact on a patient's chance of presenting with active disease and/or if this dependence is possibly not linear then it may not be appropriate to use the methodology described without investigating different functional forms for the time between visits. If however we discover that the time between visits has little or no impact on a patient's chance of presenting with active disease then it might be possible to use the approach described to model clinic visits, without being too concerned by the differing times between visits, adjusting the models by including value of $t'_{ip} - (t'_{ip-1})$ as shown in equation (4.3).

It should be noted that no interpretation can be made of the parameter α_1 in equation (4.3) when the times between visits differ between patients and are included in the models. However given further restrictions that will be placed on the models that are discussed in section 4.3 this is of no relevance in this analysis.

4.3 Analysis of The Lupus Data Using Logistic Regression Models

As previously discussed, observation of patients with lupus suggests that subsets exist within the disease. This analysis can therefore be regarded as an examination of the hypothesis 'All lupus patients present with similar patterns of disease'. Common subgroups of lupus appear to be, patients with renal disease and little else, patients with mucocutaneous and musculoskeletal disease and little else, and patients with more complicated disease. For this reason analysis has been restricted to the mucocutaneous, musculoskeletal and renal organs/systems. The mucocutaneous and musculoskeletal systems are also the most commonly affected, and the severity of renal disease leads to a particular interest in renal involvement. The methodology developed could subsequently be used to answer more complex questions about the interrelationships between disease activity in other organs/systems.

Factors that might influence the state of a patient at any visit that were included in the models as explanatory variables were; history of disease activity in the organ/system under consideration, history of disease activity in all other organs/systems, time since the last visit, time since the patient's first recorded visit, time since disease onset, where appropriate the time since the last visit at which the patient presented with disease activity in the organ/system under consideration and time since the last visit at which the patient presented with disease activity in any organ/system. Damage in the organs/systems under consideration, total damage (as measured by the SLICC index), age, sex, and age at onset of disease were also included.

Explanatory variables reflecting treatments a patient is receiving will be added into all final models in order to establish whether any relationships

found are modified by these additions.

Initial analyses showed that the most significant factor in predicting the current level of disease activity in any of the three organs/systems chosen for detailed analysis was the level of disease activity in that organ/system at the previous visit. Patients have a significantly increased chance of a BILAG

Table 4.1: % of visits with active disease (A or B score) with details of disease activity at the previous visit.

Mucocutaneous Organ/System	
Previous Visit	%
C,D or E	$\frac{633}{6454} = 9.81\%$
A or B	$\frac{292}{873} = 33.45\%$

Musculoskeletal Organ/System	
Previous Visit	%
C,D or E	$\frac{775}{6149} = 12.60\%$
A or B	$\frac{446}{1177} = 37.89\%$

Renal Organ/System	
Previous Visit	%
C,D or E	$\frac{311}{6491} = 4.79\%$
A or B	$\frac{533}{835} = 63.83\%$

score of A or B in an organ/system if they have scored either an A or B in that organ/system at the previous visit. The % of visits with active disease and the BILAG score at the previous visit for the three organs/systems chosen for analysis are given in table 4.1. Given this, and, that the primary medical focus is to identify those factors that affect a patients chance of developing active disease, the models were initially based on a subset of the visits where

the patient was known not to have any disease activity in any of the three organs/systems at the previous visit.

Table 4.2: Number of weeks since a patient's previous clinic visit with the % of patients with active disease at the previous visit.

Number of weeks since previous visit	1	2	3	4	5
% of patients who had active disease in any organ/system at the previous visit	47.6	67.9	65.6	50	50

Number of weeks since previous visit	6	7	8	9	10
% of patients who had active disease in any organ/system at the previous visit	36.1	43.2	38.7	36.3	34.9

Number of weeks since previous visit	11	12	13	14	≥ 15
% of patients who had active disease in any organ/system at the previous visit	29.1	26.4	29.2	29.0	≈ 25.0

However, table 4.2 shows the relationship between the time since the last visit and the % of patients who were observed to have active disease in any organ/system at the previous visit and it is clear that patients are more likely to return to the clinic quickly if they had active disease (in any organ/system) at the previous visit. Consequently all models were based on a subset of the visits where the patient had not only had no disease activity at their previous visit in any of the three organs/systems under consideration but also no disease activity at the previous visit in any of the other organs/systems.

In starting from a situation where the patient is known to have presented with inactive disease in all organs/systems at the previous visit there are a number of possible states the patient might present with at the next clinic visit. For the purpose of this analysis five categories are defined. They are active disease in one of the three organs/systems only, simultaneous active

disease in the mucocutaneous and musculoskeletal systems with inactive renal disease, and finally active renal disease with either active mucocutaneous disease or active musculoskeletal disease or both. The number of visits involved are given in table 4.3.

Table 4.3: Number of visits where the patient presented with active disease in one or more organs/systems having had inactive disease in all systems at their previous visit.

Inactive renal mucocutaneous and musculoskeletal disease	$\frac{2027}{2831} = 71.60\%$
Active mucocutaneous disease only	$\frac{233}{2831} = 8.23\%$
Active musculoskeletal disease only	$\frac{355}{2831} = 12.54\%$
Active renal disease only	$\frac{124}{2831} = 4.38\%$
Active mucocutaneous and musculoskeletal disease	$\frac{64}{2831} = 2.26\%$
Active renal disease with either active mucocutaneous, musculoskeletal or both	$\frac{28}{2831} = 0.99\%$

This number of categories would generally lead to the use of a polychotomous logistic regression model. However the available computer packages for polychotomous logistic regression do not permit selective inclusion of a variable in some but not all the regressions. For this reason as well as for general analytic simplicity and flexibility, especially regarding variable selection, each category was individually compared to the baseline category (of inactive renal, mucocutaneous and musculoskeletal disease) using simple logistic models. The parameters thus estimated were then combined as if they had been estimated using the full logistic model. The use of individualized logistic regression analyses as a replacement for polychotomous regression has been studied and the asymptotic relative efficiencies of the individual parameter estimates are observed to be generally high as are the efficiencies of the predicted probability estimates (Begg and Grey 1984). This approach

was also used by Allen and Farewell (2002) in the analysis of the choice of conjunction in Matthew's gospel.

No restrictions were placed on the occurrence of active disease in the five organs/systems not chosen for analysis other than the initial restriction that there was no active disease at the previous visit.

Given the particular interest in the interrelationships between disease activity in the various organs/systems, history of disease activity was included in the models in a variety of ways. Preliminary analyses indicated that the effect of a history of A scores appears to be different from the effect of a history of B scores, so separate variables were used for these for the three organs/systems chosen for detailed analysis. Binary variables were used to indicate simply whether the patient was known to have previously had the score of interest in each of the organs/systems under consideration. Dynamic covariates indicating the number of times that the patient has been observed with that score in each of the organs/systems were used to attempt to model the severity of past disease activity. (A patient is only considered to have scored an A if they had been observed to score a B,C,D or E at the previous visit, and is only considered to have scored a B if they had been observed to score a C,D or E at the previous visit.) Area under the curve measures were tried in order to attempt to reflect the effect of disease over time and the level of disease activity, and finally the number of previous visits was used as it was felt that this might reflect not only time, but the severity of the disease activity. (However it should be remembered that the number of visits would reflect disease activity in all organ systems not just the one being analysed.)

Although it has been decided to focus on the relationships between disease activity in the three organs/systems chosen for detailed analysis it is clearly necessary to include some history of disease activity in the other or-

gans/systems in any models. A single binary variable was created reflecting whether or not a patient was known to have had active disease in any of the other organs/systems. No variable was created to reflect the severity of past disease in these organs/systems; it would be meaningless to attempt to count up the number of times a patient had scored either an A or a B in five different organs/systems. Damage in each of the mucocutaneous, musculoskeletal and renal organs/systems was represented by binary variables simply indicating whether the patient is known to have damage in that organ/system. No attempt was made to represent severity of damage by using dynamic covariates to represent increases in damage (as measured by the SLICC index), as damage is only measured intermittently. Total damage was represented by a binary variable indicating whether the patient was known to have damage in any of the organs/systems. (Dynamic covariates were used for the time since the last visit at which the patient presented with disease activity in the organ/system under consideration, and the binary variables used for damage and activity can be viewed as dichotomised dynamic covariates).

As with the choice of modelling approach the choices made with regard to the variables defined are all essentially pragmatic.

When considering the effects of having histories of mucocutaneous A, mucocutaneous B, renal A, renal B, musculoskeletal A and musculoskeletal B scores it was observed that the results were being affected by a few abnormal observations. The number of visits involved are given in table 4.4.

Consequently when creating the dynamic variables for use in all the analyses counts of 4 or more mucocutaneous A's were grouped together, and coded as 4, and counts of 8 or more mucocutaneous B's were grouped together, and coded as 8. Counts of 3 or more renal A's were grouped together, and coded as 3 and counts of 6 or more renal B's were grouped together and coded as

Table 4.4: Number of visits with active disease.

	muco A	muco B	renal A	renal B	muscularA	musculoskeletal B
0	3242	2156	3340	2905	3166	1663
1	267	684	205	417	388	872
2	60	325	49	99	56	450
3	21	198	29	109	59	272
4	35	111	0	25	16	139
5	2	60	12	51		98
6	6	39		15		47
7	2	25		7		15
8		16		5		21
9		2		1		22
10		6		0		9
11		5		1		5
12		8				14
13						1
14						7

6. Counts 10 or more musculoskeletal B's were grouped together and coded as 10.

Preliminary analyses indicated that time since disease onset, time since the patient presented with disease activity in any organ/system, age, sex, age at disease onset and total damage did not significantly affect a patient's chance of presenting at a clinic visit with active disease in any of the three organs/systems chosen for detailed analysis. No details of analyses involving these variables will be given. It also appears that for the mucocutaneous, musculoskeletal and renal organs/systems a combination of the binary variable indicating whether a patient has previously scored an A, the binary variable representing whether the patient has previously scored a B and the

dynamic variables indicating the number of A's or B's the patient has scored adequately represents history of disease activity. Consequently results are not given for area under the curve measures, and the variable representing the number of visits.

The validity of the linear assumption for all variables used was assessed by fitting generalized additive models (Hastie and Tibshirani 1986). Generalized additive models provide a method for identifying non-linear covariate effects in exponential family models.

In the linear logistic models proposed for the analysis of the lupus data it is assumed that the covariates act in a linear fashion and that

$$\text{logit} \Pr\left(y_{it} = 1 \mid y_{it-1}\right) = \beta_0 + \beta_1 X_1 + ... + \beta_p X_p.$$

The generalized additive logistic model assumes instead that

$$\text{logit} \Pr\left(y_{it} = 1 \mid y_{it-1}\right) = s_0 + \sum_{1}^{p} s_i(X_i)$$

where $s_1(.), ..., s_p(.)$ are unspecified nonparametric functions. These functions are estimated in a flexible manner and can reveal possible non linearities in the effects of the covariates.

Consider a scatterplot of points (x_i, y_i), where y is the response variable and x is a prognostic factor. It is desirable to fit a smooth curve to $s(x)$ that summarizes the dependence of y on x. However the curve $s(x)$ that minimizes $\sum(y_i - f(x_i))^2$ is generally an interpolating curve that is not smooth at all. To overcome this the estimation of the functions $s_1(.), ..., s_p(.)$ was done using a cubic spline smoother. The cubic spline smoother imposes smoothness on $s(x)$. It minimizes

$$\sum[y_i - s(x_i)]^2 + \lambda \int s''(x)^2 dx,$$

91

where λ is a non negative smoothing parameter that governs the trade off between the 'goodness of fit' to the data and the shape of $s(x)$. Larger values of λ force $s(x)$ to be smoother. For the analysis undertaken $\lambda = 3$.

Plots from generalized additive models were used to identify covariates that might not act in a linear fashion. Once the covariates had been identified, quadratic and higher order terms were added to the linear logistic models and fitted using maximum likelihood techniques. All final model selection was based on this fully parametric fitting. Where appropriate quadratic and/or higher order factors have been included in the models. An example of the plot for the number of musculoskeletal scores, in the model for active musculoskeletal disease, is given in figure 4.1.

For each final model, disease history was represented by a combination of those variables that significantly affected the fit of the model. A 5% significance level was used as a guideline in all analyses. Initial variable selection was done using generalized linear models. Only the final models in each section were fit using generalized estimating equations with an exchangeable correlation structure.

Figure 4.1: Plot of generalized additive model used to assess the linearity of the components
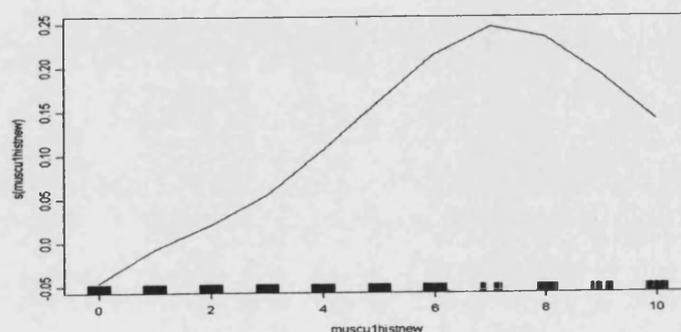
The goodness of fit of the final models is assessed using a method proposed

92

by Horton *et al.* (1999), and the final models were validated using a second data set.

4.3.1 Assessing the Goodness of Fit of the Models.

Following the suggestion of Hosmer and Lemeshow (1989) for ordinary logistic regression Horton *et al.* (1999) propose forming G (usually 10) groups based on combinations of the covariates x_{it}'s in the logistic regression model, and testing to see if additional regression coefficients from G-1 indicator variables (representing the groups formed) differ significantly from 0.

It is however necessary to have a rule for forming the groups based on combinations of the covariates x_{it}. If all the covariates are discrete, different groups can be formed for each level in the cross-classification of covariates, but with many discrete covariates there are too many groups. With many discrete and/or continuous covariates Horton *et al.* (1999) suggest forming groups based on deciles of risk, as suggested by Hosmer and Lemeshow (1989). That is the groups are formed based on

$$\hat{p}_{it} = \frac{e^{\hat{\alpha} + \hat{\beta} x_{it}}}{1 + e^{\hat{\alpha} + \hat{\beta} x_{it}}}.$$

Each subject has T_i separate estimates of risk ($\hat{p}_{it}$'s), and there are $\sum_{i=1}^{N} T_i$ observations in total. Horton *et al.* (1999) suggest forming 10 groups of approximately equal size in the following manner:

1. The first group contains the $\sum_{i=1}^{N} T_i / 10$, (y_{it}, x_{it})'s with the smallest values of $\hat{p}_{it}$.

2. The second group contains the $\sum_{i=1}^{N} T_i / 10$, (y_{it}, x_{it})'s with the next smallest values of $\hat{p}_{it}$.

and so on.

10. The last group contains the $\sum_{i=1}^{N} T_i/10$, $(y_{it}, \boldsymbol{x}_{it})$'s with the largest values of $\hat{p}_{it}$.

Because some subjects may have identical covariate values, there can be ties in predicted risk, and so the number of subjects in each decile of risk may differ slightly.

In general G groups could be formed with approximately $\sum_{i=1}^{N} T_i/G$ observations in each group. Since subject i can have different $\hat{p}_{it}$'s for each of the T_i observations, a subject's group membership, g, can change for different t, $g = 1, ..., G - 1$. That is a group variable can be considered to be a time-varying covariate. Observations in the same group have similar $\hat{p}_{it}$'s and thus similar predicted risks.

Suppose we define the $(G - 1)$ group indicators

$$I_{itg} = 1 \quad \text{if } \hat{p}_{it} \text{ is in group g, 0 otherwise,}$$

$g = 1, ..., G - 1$, where the groups are based on percentiles of risk. Then to test the goodness-of-fit of the model

$$\text{logit}(p_{it}) = \alpha + \beta \boldsymbol{x}_{it}$$

the alternative model

$$\text{logit}(p_{it}) = \alpha + \beta \boldsymbol{x}_{it} + \gamma_1 I_{it1} + ... + \gamma_{G-1} I_{it,G-1}.$$

is considered. Effectively an 'alternative' model is being used to test the fit of the given model. Moore and Spruill (1975) showed that, asymptotically, the partition can be treated as though based on the true p_{it}. Thus I_{itg} can be treated as a 'fixed' covariate. Results of simulations suggest that this assumption holds for moderately sized samples.

94

In general if the model

$$\text{logit}(p_{it}) = \alpha + \beta \boldsymbol{x}_{it}$$

is appropriate then $\gamma_1, ..., \gamma_{G-1} = 0$. A test of the fit of the model is equivalent to the test of:

$$H_0 : \gamma_1 = ... = \gamma_{G-1} = 0$$

which will be implemented using global Wald statistics.

Chapter 5

Results

A number of analyses have been undertaken. Table 5.1 sets out the order of the analyses.

Table 5.1: Order of Analyses

Models	Results	Goodness of fit
Simple logistic regression models, for the five outcomes, detailed on p86, used to estimate parameters of the full polychotomous logistic model	Sections 5.1.1-5.1.5	Section 5.1.6
Simple logistic regression models, for the three outcomes active mucocutaneous, active renal, and active musculoskeletal disease	Section 5.2.1	Section 5.2.2
Simple logistic regression models, for the three outcomes active mucocutaneous, active renal, and active musculoskeletal disease with medication included	Section 5.2.4	NA
The final models are validated on a second data set	Section 5.3.1	NA

Details on the number of patients with different combinations of disease

activity are given in table 5.2.

In this chapter only the main findings of the analyses are presented. The final models in each section have been fitted using the generalized estimating equation approach of Liang and Zeger (1986) as discussed in section 4.2.1. The results from the full analyses are given in appendix D. This includes the analysis of each outcome by individual organ/system and the selection of the variables that best represents history of activity in that organ/system. A discussion of how the final multivariate models were selected is also given, with details of combinations of variables that were considered, discussion of why some effects found in the univariate analyses disappear, and the full process of variable selection.

All odds ratios given for the number of times a patient has been observed to score BILAG A's or B's are the odds ratios associated with each additional observation of a BILAG A or B.

5.1 Results from the Simple Logistic Regression Models Used to Estimate the Parameters of the Full Logistic Model.

5.1.1 The Mucocutaneous Organ/System

Univariate Analyses

Initial univariate analyses were carried out in order to assess the relationships between history of disease activity in the organs/systems chosen for detailed analysis and a patient's chance of presenting at a clinic visit with active mucocutaneous disease. The results are given, with unadjusted odds ratios

Table 5.2: Numbers of patients with histories of disease activity

Mucocutaneous B	Mucocutaneous A	Musculo skeletal B	Musculo skeletal A	Renal B	Renal A	Other involvement	No other involvement
						19	24
					x	1	
				x		2	4
			x			2	
			x	x	x	1	
		x				35	22
		x			x	1	
		x		x		10	2
		x		x	x	3	
		x	x			4	3
		x	x	x		1	1
x						18	12
x					x	1	
x				x		4	2
x				x	x	3	
x			x	x	x	1	
x		x				51	14
x		x			x	1	
x		x		x		20	1
x		x		x	x	10	
x		x	x			14	2
x		x	x	x		8	
x		x	x	x	x	4	
x	x					8	2
x	x			x		2	1
x	x			x	x	2	
x	x		x	x		1	
x	x	x				18	6
x	x	x		x		11	1
x	x	x	x			4	1
x	x	x	x	x		4	
x	x	x	x	x	x	5	

Table 5.3: Coefficients with unadjusted odds ratios and p values from the univariate analyses of active mucocutaneous disease vs. inactive mucocutaneous, musculoskeletal and renal disease.

	coefficient	odds ratio	p value
A previous occurrence of a mucocutaneous B	1.102	3.010	≈ 0
Number of mucocutaneous B scores	0.547		
(Number of mucocutaneous B scores)^2	-0. 048	$e^{0.499-0.096mb*}$	≈ 0
A previous occurrence of a mucocutaneous A	1.225	3.404	≈ 0
Number of mucocutaneous A scores	1.109		
(Number of mucocutaneous A scores)^2	-0.173	$e^{0.936-0.346ma*}$	≈ 0
A previous occurrence of a renal B	-0.122	0.885	0.499
Number of renal B scores	-0.086	0.917	0.291
A previous occurrence of a renal A	-0.828	0.437	0.024
Number of renal A scores	-1.891		
(Number of renal A scores)^2	0.672	$e^{-1.219+1.344ra*}$	0.006
A previous occurrence of a musculoskeletal B	-0.020	0.980	0.884
Number of musculoskeletal B scores	0.048	1.050	0.171
A previous occurrence of a musculoskeletal A	-0.296	0.743	0.202
Number of musculoskeletal A scores	-0.130	0.878	0.369
Mucocutaneous damage	1.195	3.304	≈ 0
Renal damage	-0.778	0.459	0.255
Musculoskeletal damage	-0.519	0.595	0.052
History of disease activity in all other organs/systems	0.123	1.131	0.426
Time (years) since first clinic visit	0.024	1.025	0.331
Time (years) since previous clinic visit	-0.368	0.692	0.289
Time (years) since last visit with active mucocutaneous disease	-0.028	0.973	0.492

* Where mb, ma, and ra are the number of mucocutaneous Bs, mucocutaneous As and renal As respectively that the comparison patient had been observed to score at the previous visit.

and p values in table 5.3.

A history of either mucocutaneous A scores or mucocutaneous B scores has the most noticeable effect on a patient's chance of presenting with active mucocutaneous disease; in both cases the patient's chance is significantly increased.

A history of renal A scores generally appears to significantly decrease a patient's chance of presenting with active mucocutaneous disease, as indicated by the negative coefficient of a previous occurrence of a renal A, and the negative coefficient of the number of renal A scores. There is however some evidence of non linearity in this univariate analysis.

A history of any of renal B, musculoskeletal A or musculoskeletal B scores does not significantly affect a patient's chance of presenting with active mucocutaneous disease.

Mucocutaneous damage significantly increases a patient's chance of presenting with active disease. Neither renal nor musculoskeletal damage has a significant effect.

A history of disease activity in the organs/systems not chosen for detailed analysis does not significantly affect a patient's chance of presenting with active mucocutaneous disease.

Time since the fist clinic visit, time since the previous visit and the time since the patient was last observed with active mucocutaneous disease all have no significant effect on a patient's chance of presenting with active mucocutaneous disease.

Multivariate Analysis

All variables were now included in a single model with only main effects. Variables that in this analysis did not significantly affect a patient's chance

of presenting with active mucocutaneous disease were removed, giving the results shown with adjusted odds ratios and p values in table 5.4.

Table 5.4: Coefficients with adjusted odds ratios and p values from the multivariate analysis (with only main effects) of active mucocutaneous disease vs. inactive mucocutaneous, musculoskeletal and renal disease.

	coefficient	odds ratio	p value
A previous occurrence of a mucocutaneous B	0.894	2.446	≈ 0
Number of mucocutaneous B scores	0.187	1.206	0.001
A previous occurrence of a mucocutaneous A	1.106	3.021	≈ 0
A previous occurrence of a renal A	-1.098	0.334	0.003
A previous occurrence of a musculoskeletal A	-0.599	0.549	0.020
Time (years) since first clinic visit	-0.140	0.869	0.00003

As was indicated by the univariate analyses a history of mucocutaneous B scores and a history of mucocutaneous A scores increase a patient's chance of presenting with active mucocutaneous disease; the chance of presenting with active mucocutaneous disease increasing further with each additional observation of a mucocutaneous B.

A history of either renal A or musculoskeletal A scores decreases a patient's chance of presenting with active mucocutaneous disease.

A patient's chance of presenting with active mucocutaneous disease decreases with the time that they have been registered at the clinic.

Interactions

An investigation into interactions was then undertaken. Significant interactions were found between a previous occurrence of a mucocutaneous A and a previous occurrence of a musculoskeletal A, between a previous occurrence

of a mucocutaneous A and a previous occurrence of a musculoskeletal B and between a previous occurrence of a musculoskeletal B and a previous occurrence of a renal B. All other variables that were significant in the previous multivariate analysis have remained significant and those that were not have remained so. The results of this analysis (fitted using generalized estimating equations) with adjusted odds ratios and 95% confidence intervals are given in table 5.5.

Table 5.5: Coefficients with adjusted odds ratios (with 95% CI) from the final multivariate analysis of active mucocutaneous disease vs. inactive mucocutaneous, musculoskeletal and renal disease.

	coefficient	odds ratio
A previous occurrence of a mucocutaneous B	0.774	2.169 [1.426,3.300]
Number of mucocutaneous B scores	0.191	1.211 [1.087,1.348]
A previous occurrence of a renal A	-0.858	0.424 [0.198,0.910]
Time (years) since first clinic visit	-0.137	0.872 [0.816,0.932]
A previous occurrence of a mucocutaneous A	2.080	-
A previous occurrence of a musculoskeletal B	-0.056	-
A previous occurrence of a musculoskeletal A	-0.206	-
A previous occurrence of a renal B	-1.100	-
A previous occurrence of a mucocutaneous A* A previous occurrence of a musculoskeletal B	-1.296	-
A previous occurrence of a mucocutaneous A* A previous occurrence of a musculoskeletal A	-1.296	-
A previous occurrence of a renal B* A previous occurrence of a musculoskeletal B	1.230	-

A history of mucocutaneous A scores significantly increases a patient's chance of presenting with active mucocutaneous disease. However the interactions between a previous occurrence of a musculoskeletal A and a previous

occurrence of a mucocutaneous A, and between a previous occurrence of a musculoskeletal B and a previous occurrence of a mucocutaneous A indicate that for those patients with a history of either musculoskeletal A or B scores the effect of a history of mucocutaneous A scores is reduced.

Although the analysis presented in section 5.1.1, suggests that a history of musculoskeletal A scores decreases a patient's chance of presenting with active mucocutaneous disease it now appears that this effect is only significant for those patients with a history of mucocutaneous A scores. Neither a history of musculoskeletal A scores nor a history of musculoskeletal B scores has a significant effect on the chance of a patient with no history of mucocutaneous A scores presenting with active mucocutaneous disease.

Table 5.6: Odds ratios* for combinations of a history of mucocutaneous A scores, a history of musculoskeletal A scores, a history of musculoskeletal B scores and a history of renal B scores.

	No previous occurrence of a mucocutaneous A	Previous occurrence of a mucocutaneous A
No history of musculoskeletal A, musculoskeletal B or renal B scores	1. 000*	8. 003 [4.100,15.024]
A history of musculoskeletal A scores	0.814 [0.488,1.359]	1.782 [0.545,5.824]
A history of musculoskeletal B scores	0.945 [0.644,1.388]	2.071 [1.101,3.896]
A history of renal B scores	0.333 [0.140,0.792]	2.663 [1.236,5.738]
A history of musculoskeletal B and renal B scores	1.076 [0.595,1.946]	2.357 [1.094,5.078]
A history of musculoskeletal A and musculoskeletal B scores	0.770 [0.426,1.392]	0. 461 [0.164,1.299]

* All odds rations are compared to a common default category

A history of renal B scores decreases a patient's chance of presenting with

active mucocutaneous disease. However the significant interaction between a history of musculoskeletal B scores and a history of renal B scores suggests that a history of renal B scores only affects those patients with no history of musculoskeletal B scores.

All odds ratios involving the interactions are given in table 5.6. The confidence intervals for the odds ratios were calculated using the correlation of the coefficients from the generalized linear model and the robust variance-covariance matrix from the model fitted using generalized estimating equations.

5.1.2   The Renal Organ/System

Univariate Analyses

Initial univariate analyses were carried out in order to assess the relationships between history of disease activity in the organs/systems chosen for detailed analysis and a patient's chance of presenting with active renal disease at a clinic visit. The results are given with unadjusted odds ratios and p values in table 5.7.

A history of either renal A scores or renal B scores has the most noticeable effect on a patient's chance of presenting with active renal disease; in both cases the patient's chance is significantly increased.

A history of either mucocutaneous A or mucocutaneous B scores significantly increases a patient's chance of presenting with active renal disease.

Neither a history of musculoskeletal A nor musculoskeletal B scores significantly affects a patient's chance of presenting with active renal disease.

Both renal and mucocutaneous damage significantly increase a patient's chance of presenting with active renal disease. Musculoskeletal damage does

Table 5.7: Coefficients with unadjusted odds ratios and p values from the univariate analyses of active renal disease vs. inactive mucocutaneous, musculoskeletal and renal disease.

	coefficient	odds ratio	p value
A previous occurrence of a renal B	2.456	11.654	≈ 0
Number of renal B scores	0.543		
(Number of renal B scores)^2	-0.162	$e^{1.290-0.324rb*}$	≈ 0
A previous occurrence of a renal A	2.216	9.173	≈ 0
Number of renal A scores	2.149		
(Number of renal A scores)^2	-0.389	$e^{1.759-0.778ra*}$	≈ 0
A previous occurrence of a musculoskeletal B	-0.209	0.811	0.255
Number of musculoskeletal B scores	-0.096	0.908	0.128
A previous occurrence of a musculoskeletal A	0.175	1.192	0.511
Number of musculoskeletal A scores	0.071	1.074	0.664
A previous occurrence of a mucocutaneous B	0.381	1.464	0.038
Number of mucocutaneous B scores	0.116	1.123	0.025
A previous occurrence of a mucocutaneous A	0.613	1.847	0.015
Number of mucocutaneous A scores	0.295	1.343	0.015
Renal damage	2.745	15.565	≈ 0
Musculoskeletal damage	-0.070	0.932	0.822
Mucocutaneous damage	0.990	2.460	0.0004
History of disease activity in all other organs/systems	0.274	1.315	0.209
Time (years) since first clinic visit	0.064	1.066	0.049
Time (years) since previous clinic visit	-0.290	0.748	0.516
Time (years) since last visit with active renal disease	0.221	1.247	0.00004

* Where rb and ra are the number of renal B's and renal A's respectively that the comparison patient had been observed to score at the previous visit.

not have a significant effect.

History of disease activity in the organs/systems not chosen for detailed

analysis does not significantly affect a patient's chance of presenting with active renal disease.

Time since the first clinic visit and time since the last clinic visit where the patient was observed to have active renal disease both significantly increase a patient's chance of presenting with active renal disease. The time since the previous clinic visit has no significant effect on a patient's chance of presenting with active renal disease.

Multivariate Analysis

All variables were now included in a single model with only main effects. Variables that in this analysis did not significantly affect a patient's chance of presenting with active renal disease were removed, giving the results shown with adjusted odds ratios and p values in table 5.8.

Table 5.8: Coefficients with adjusted odds ratios and p values from the multivariate analysis (with only main effects) of active renal disease vs. inactive mucocutaneous, musculoskeletal and renal disease.

	coefficient	odds ratio	p value
Previous occurrence of a renal B	1.578	4.846	≈ 0
Number of renal B scores	0.272	1.312	0.002
Previous occurrence of a renal A	0.732	2.032	0.004
Number of musculoskeletal B scores	-0.244	0.709	0.001
Renal damage	1.247	3.478	0.00004

As was indicated by the univariate analyses a history of renal B scores increases a patient's chance of presenting with active renal disease, the chance increasing further with each additional observation of a renal B.

106

A history of renal A scores increases a patient's chance of presenting with active renal disease.

A history of musculoskeletal B scores decreases a patient's chance of presenting with active renal disease, the chance decreasing further with each additional observation of a musculoskeletal B.

Renal damage increases a patient's chance of presenting with active renal disease.

Interactions

An investigation into interactions was subsequently undertaken. A significant interaction was found between a previous occurrence of a renal B and a previous occurrence of a renal A. All variables that were found to be significant in the previous multivariate analysis have remained significant, and those that were not have remained so. The results of this analysis with adjusted odds ratios are shown in table 5.9.

Table 5.9: Coefficients with adjusted odds ratios and p values from the multivariate analysis of active renal disease vs. inactive mucocutaneous, musculoskeletal and renal disease.

	Coefficient	odds ratio	p value
Number of renal B scores	0.303	1.353	0.0004
Number of musculoskeletal B scores	-0.249	0.779	0.0009
Renal damage	1.287	3.624	0.00002
A previous occurrence of a renal B	1.873	-	≈ 0
A previous occurrence of a renal A	2.918	-	≈ 0
A previous occurrence of a renal B*			
A previous occurrence of a renal A	-2.526	-	≈ 0

The interaction between a previous occurrence of a renal B and a previous occurrence of a renal A indicates that they are not additive in effect. If a patient has a history of both renal A and renal B scores then the effect of a history of renal A scores is almost negated. The numbers of patients involved are given in table 5.10 and the odds ratios are given in table 5.11.

Table 5.10: Numbers of patients with histories of renal A and B scores.

	No history of renal A scores	History of renal A scores
No history of renal B scores	259	4
History of renal B scores	73	31

Table 5.11: Odds ratios for combinations of a previous occurrence of a renal B and a previous occurrence of a mucocutaneous A.

	No previous occurrence of a renal B score	previous occurrence of a renal B score
No previous occurrence of a renal A score	1.000	6.508
Previous occurrence of a renal A score	18.508	9.632

Given this, a new variable was created representing a history of renal A scores only (that is a history of renal A scores, but no history of renal B scores).

Final Model

Replacing a previous occurrence of a renal A, and the interaction between a previous occurrence of a renal A and a previous occurrence of a renal B

with a single binary variable representing a history of renal A scores only, led to a change in residual deviance of 2.253 which is not significant on χ_1^2 (p= 0.133). The results of this analysis (fitted using generalized estimating equations) are given with adjusted odds ratios and 95% confidence intervals in table 5.12.

Table 5.12: Coefficients with adjusted odds ratios from the final multivariate analysis of active renal disease vs. inactive mucocutaneous, musculoskeletal and renal disease.

	coefficient	odds ratio
A previous occurrence of a renal B	1.852	6.371 [3.155,12.867]
A previous occurrence of a renal A only	2.909	18.335 [6.042,55.643]
Number of renal B scores	0.339	1.404 [1.175,1.677]
Number of musculoskeletal B scores	-0.237	0.789 [0.676,0.921]
Renal damage	1.324	3.759 [2.277,6.204]

A history of renal B scores increases a patient's chance of presenting with active renal disease, the chance increasing further with each additional observation of a renal B. This effect is the same for those patients who also have a history of renal A scores.

Patients with a history of renal A scores only also have an increased chance of presenting with active renal disease. Their chance however does not increase further with an additional observation of a renal A.

Each additional observation of a musculoskeletal B decreases a patient's chance of presenting with active renal disease.

Renal damage increases a patient's chance of presenting with active renal disease.

5.1.3 The Musculoskeletal Organ/System

Univariate Analyses

Initial univariate analyses were carried out in order to assess the relationships between history of disease activity in the organs/systems chosen for detailed analysis and a patient's chance of presenting at a clinic visit with active musculoskeletal disease. The results are given with unadjusted odds ratios and p values in table 5.13.

A history of either musculoskeletal A scores or musculoskeletal B scores has the most noticeable effect on a patient's chance of presenting with active musculoskeletal disease; in both cases the patient's chance is significantly increased.

A history of renal B scores significantly decreases a patient's chance of presenting with active musculoskeletal disease.

A history of mucocutaneous B scores significantly increases a patient's chance of presenting with active musculoskeletal disease. Neither a history of renal A scores nor a history of mucocutaneous A scores significantly affects a patient's chance of presenting with active musculoskeletal disease.

Musculoskeletal damage significantly increases a patient's chance of presenting with active musculoskeletal disease. Neither renal nor mucocutaneous damage have any significant effect.

A history of disease activity in the organs/systems not chosen for detailed analysis does not significantly affect a patient's chance of presenting with active musculoskeletal disease.

A patient's chance of presenting with active musculoskeletal disease decreases with the time since the last clinic visit at which the patient was observed to have active musculoskeletal disease.

Table 5.13: Coefficients with unadjusted odds ratios and p values from the univariate analyses of active musculoskeletal disease vs. inactive mucocutaneous, musculoskeletal and renal disease.

	coefficient	odds ratio	p value
A previous occurrence of a musculoskeletal B	0.912	2.420	≈ 0
Number of musculoskeletal B scores	0.492		
(Number of musculoskeletal B scores)^2	-0.033	$e^{0.459-0.066*m_u b*}$	≈ 0
A previous occurrence of a musculoskeletal A	0.844	2.325	≈ 0
Number of musculoskeletal A scores	0.501	1.650	≈ 0
A previous occurrence of a renal B	-0.280	0.756	0.075
Number of renal B scores	-0.472		
(Number of renal B scores)^2	0.108	$e^{-0.364+0.216*rb*}$	0.007
A previous occurrence of a renal A	-0.481	0.618	0.061
Number of renal A scores	-0.245	0.783	0.150
A previous occurrence of a mucocutaneous B	0.255	1.291	0.025
Number of mucocutaneous B scores	0.065	1.608	0.060
A previous occurrence of a mucocutaneous A	-0.085	0.919	0.670
Number of mucocutaneous A scores	0.090	1.094	0.342
Musculoskeletal damage	0.786	2.195	≈ 0
Renal damage	0.418	1.518	0.232
Mucocutaneous damage	-0.091	0.913	0.727
History of disease activity in all other organs/systems	-0.104	0.901	0.399
Time (years) since first clinic visit	0.034	1.035	0.097
Time (years) since previous clinic visit	0.148	1.160	0.362
Time (years) since last visit with active musculoskeletal disease	-0.106	0.899	0.008

* Where $m_u b$, and rb are the number of musculoskeletal Bs, and renal Bs respectively that the comparison patient had been observed to score at the previous visit.

111

Time since the first clinic visit and the time since the previous clinic visit do not significantly affect a patient's chance of presenting with active musculoskeletal disease.

Multivariate Analysis

All variables were now included in a single model with only main effects. Variables that in this analysis did not significantly affect a patient's chance of presenting with active musculoskeletal disease were removed, giving the results shown with adjusted odds ratios and p values in table 5.14.

Table 5.14: Coefficients with adjusted odds ratios and p values from the multivariate analysis (with only main effects) of active musculoskeletal disease vs. inactive mucocutaneous, musculoskeletal and renal disease.

	coefficient	odds ratio	p value
Number of musculoskeletal B scores	0.695		
(Number of musculoskeletal B scores)^2	-0.050	$e^{0.645-0.100m_u b}$	≈ 0
A previous occurrence of a musculoskeletal A	0.341	1.406	0.032
A previous occurrence of a renal B	-0.523	0.593	0.002
History of disease activity in all other organs/systems	-0.419	0.658	0.003
Time (years) since first clinic visit	-0.066	0.936	0.032
Time (years) since previous clinic visit	0.403	1.497	0.010
Time (years) since last visit with active musculoskeletal disease	-0.172	0.842	0.002

* Where $m_u b$ is the number of musculoskeletal Bs, that the comparison patient had been observed to score at the previous visit.

As was indicated by the univariate analyses, a history of musculoskeletal B scores increases a patient's chance of presenting with active musculoskeletal disease, the chance increasing further with each additional observation of

a musculoskeletal B. However a quadratic factor is needed to adequately represent the number of musculoskeletal B scores. The effect of an additional musculoskeletal B decreases with the number of musculoskeletal Bs a patient has been observed to score.

A history of musculoskeletal A scores also increases a patient's chance of presenting with active musculoskeletal disease.

A history of renal B scores and a history of disease activity in the five organs/systems not chosen for detailed analysis both decrease a patient's chance of presenting with active musculoskeletal disease.

A patient's chance of presenting with active musculoskeletal disease decreases with the time they have been registered at the clinic, and decreases with the time since they last presented with a musculoskeletal A or B. However a patient's chance of presenting with active musculoskeletal disease increases with the time since the previous clinic visit.

Interactions

An investigation into interactions was then undertaken. Significant interactions were found between a previous occurrence of a musculoskeletal B and a previous occurrence of a musculoskeletal A and between a previous occurrence of a musculoskeletal A and a previous occurrence of a mucocutaneous A. All variables that were found to be significant in the previous multivariate analysis have remained significant, apart from time since the first clinic visit, and those that were not significant have remained so. The results of this analysis with adjusted odds ratios and p values are given in table 5.15.

The interaction between a previous occurrence of a musculoskeletal B and a previous occurrence of a musculoskeletal A indicates that the effect of the two is less than additive. If a patient has a history of both musculoskeletal A

Table 5.15: Coefficients with adjusted odds ratios and p values from the multivariate analysis of active musculoskeletal disease vs. inactive mucocutaneous, musculoskeletal and renal disease.

	Coefficient	odds ratio	p value
Number of musculoskeletal B scores	0.572		
(Number of musculoskeletal B scores)^2	-0.042	$e^{0.530-0.084m_u b*}$	≈ 0
A previous occurrence of a renal B	-0.501	0.606	0.005
History of disease scores in all other organs/systems	-0.464	0.629	0.001
Time (years) since previous clinic visit	0.353	1.423	0.025
Time (years) since last visit with active musculoskeletal disease	-0.233	0.792	0.00003
A previous occurrence of a musculoskeletal B	0.418	-	0.075
A previous occurrence of a musculoskeletal A	1.592	-	0.00009
A previous occurrence of a mucocutaneous A	-1.021	-	0.004
A previous occurrence of a musculoskeletal B* A previous occurrence of a musculoskeletal A	-1.598	-	0.0003
A previous occurrence of a musculoskeletal A* A previous occurrence of a mucocutaneous A	1.352	-	0.003

* Where $m_u b$ is the number of musculoskeletal Bs, that the comparison patient had been observed to score at the previous visit.

Table 5.16: Odds ratios for combinations of a previous occurrence of a musculoskeletal B and a previous occurrence of a musculoskeletal A.

	No previous occurrence of a musculoskeletal B score	previous occurrence of a musculoskeletal B score
No previous occurrence of a musculoskeletal A score	1.000	1.519
Previous occurrence of a musculoskeletal A score	4.192	1.509

and musculoskeletal B scores then the effect of a history of musculoskeletal A scores is effectively negated. The odds ratios are given in table 5.16 and numbers of patients involved are given in table 5.17.

Table 5.17: Numbers of patients with histories of musculoskeletal A and B scores.

	No history of musculoskeletal A scores	History of musculoskeletal A scores
No history of musculoskeletal B scores	105	5
History of musculoskeletal B scores	206	51

Given this a new variable was created representing a history of musculoskeletal A scores only (that is a history of musculoskeletal A scores, but no history of musculoskeletal B scores).

Final Model

Replacing the interaction between a previous occurrence of a musculoskeletal B and a previous occurrence of a musculoskeletal A (and a previous occurrence of a musculoskeletal B) with the variable representing a history of musculoskeletal A scores only, led to a change in deviance of 3.135 which is not significant on χ^2_1 (p= 0.077). The results of this final analysis (fitted using generalized estimating equations) are given with adjusted odds ratios and 95% confidence intervals in table 5.18.

As indicated by previous analyses a history of active musculoskeletal disease increases a patient's chance of presenting with active musculoskeletal disease. This is represented in the model by the number of times a patient has scored a musculoskeletal B (odds ratios are given in table 5.19), and by

115

Table 5.18: Coefficients with adjusted odds ratios from the final multivariate analysis of active musculoskeletal disease vs. inactive mucocutaneous, musculoskeletal and renal disease.

	coefficient	odds ratio
Number of musculoskeletal B scores	0.554	
(Number of musculoskeletal B scores)^2	-0.045	$e^{0.509-0.089m_u b*}$
A previous occurrence of a musculoskeletal A only	1.264	3.541 [1.751,7.161]
A previous occurrence of a renal B	-0.482	0.618 [0.429,0.890]
History of disease activity in all other organs/systems	-0.396	0.673 [0.478,0.950]
Time (years) since previous musculoskeletal A or B	-0.141	0.868 [0.787,0.958]
A previous occurrence of a musculoskeletal A	0.059	-
A previous occurrence of a mucocutaneous A	-1.089	-
A previous occurrence of a musculoskeletal A*		
A previous occurrence of a mucocutaneous A	1.531	-

* Where $m_u b$ is the number of musculoskeletal Bs, that the comparison patient had been observed to score at the previous visit.

a previous occurrence of a musculoskeletal A for those patients who have scored only musculoskeletal A's not B's.

A history of renal B scores and a history of disease activity in all organs/systems not chosen for detailed analysis decrease a patient's chance of presenting with active musculoskeletal disease.

The significant interaction between a previous occurrence of a musculoskeletal A and a previous occurrence of a mucocutaneous A indicates that, a history of mucocutaneous A scores increases the chance of patients with a history of musculoskeletal A scores presenting with active musculoskeletal disease. However for those patients with no history of musculoskeletal A scores the chance of presenting with active musculoskeletal disease is de-

Table 5.19: Adjusted odds ratios for each additional musculoskeletal B score.

Number of musculoskeletal B's scored by the previous visit	Odds ratios for a musculoskeletal B score
0	1.664
1	1.592
2	1.523
3	1.456
4	1.393
5	1.332
6	1.274
7	1.219
8	1.165
9	1.115

creased by a history of mucocutaneous A scores.

A patient's chance of presenting with active musculoskeletal disease decreases with the time since they last scored a musculoskeletal A or B. (The time since the previous visit loses significance when the model is fitted using generalized estimating equations.)

Odds ratios for combinations of musculoskeletal and mucocutaneous A scores are given in table 5.20.

5.1.4 Simultaneous Mucocutaneous and Musculoskeletal Disease Activity

Univariate Analyses

Initial univariate analyses were carried out in order to assess the relationships between history of disease activity in the organs/systems chosen for detailed

Table 5.20: Odds ratios for combinations of a previous occurrence of a musculoskeletal A and a previous occurrence of a mucocutaneous A.

	No Previous occurrence of a musculoskeletal A score	Previous occurrence of a musculoskeletal A score
No previous occurrence of a mucocutaneous A score	1	1.061* [0.649,1.732]
Previous occurrence of a mucocutaneous A score	0.337 [0.152,0.744]	1.651 [1.039,2.263]

*this odds ratio should be interpreted with caution as the effects of musculoskeletal A and B scores have already been adjusted for

analysis and a patient's chance of presenting at a clinic visit with active disease in both the mucocutaneous and musculoskeletal organs/systems. The results are given with unadjusted odds ratios and p values in table 5.21.

A history of musculoskeletal B or mucocutaneous B scores significantly increases a patient's chance of presenting with both active mucocutaneous and musculoskeletal disease.

Neither a history of mucocutaneous A scores nor a history of musculoskeletal A scores significantly affect a patient's chance of presenting with both active mucocutaneous and musculoskeletal disease.

A history of renal B scores does not significantly affect a patient's chance of presenting with both active mucocutaneous and musculoskeletal disease.

No patients with a history of renal A scores have ever presented at a clinic visit (with inactive disease in all organs/systems at the previous visit) with simultaneous mucocutaneous and musculoskeletal disease. As a consequence of this no further mention of a history of renal A scores will be made in this section.

Musculoskeletal damage significantly affects a patient's chance of presenting with both active mucocutaneous and musculoskeletal disease. Neither

Table 5.21: Coefficients with unadjusted odds ratios and p values from the univariate analyses of active musculoskeletal and mucocutaneous disease vs. inactive mucocutaneous, musculoskeletal and renal disease.

	coefficient	odds ratio	p value
A previous occurrence of a musculoskeletal B	0.696	2.006	0.011
Number of musculoskeletal B scores	0.153	1.165	0.002
A previous occurrence of a musculoskeletal A	0.108	1.114	0.767
Number of musculoskeletal A scores	0.122	131	0.523
A previous occurrence of a mucocutaneous B	0.721	2.057	0.005
Number of mucocutaneous B scores	0.161	1.175	0.009
A previous occurrence of a mucocutaneous A	0.197	1.218	0.606
Number of mucocutaneous A scores	0.013	1.013	0.951
A previous occurrence of a renal B	-0.429	0.651	0.258
Number of renal B scores	-0.226	0.798	0.231
A previous occurrence of a renal A	-6.297	NA	NA
Number of renal A scores	-5.446	NA	NA
Musculoskeletal damage	0.728	2.070	0.022
Mucocutaneous damage	0.059	1.061	0.910
Renal damage	0.504	1.656	0.489
History of disease activity in all other organs/systems	-0.238	0.788	0.375
Time (years) since first clinic visit	-0.108	0.897	0.040
Time (years) since previous clinic visit	-1.082	0.339	0.206
Time (years) since last visit with active musculoskeletal disease	-0.265	0.767	0.021
Time (years) since last visit with active mucocutaneous disease	-0.099	0.905	0.244

mucocutaneous nor renal damage have a significant effect.

A history of disease activity in the five organs/systems not chosen for detailed analysis does not significantly affect a patient's chance of presenting with both active mucocutaneous and musculoskeletal disease.

A patient's chance of presenting with active mucocutaneous and musculoskeletal disease decreases with the time since their first clinic visit and with the time since they were last observed to score a musculoskeletal A or B.

The time since the previous visit and the time since they were last observed to score a mucocutaneous A or B have no significant effect on a patient's chance of presenting with active mucocutaneous and musculoskeletal disease.

Multivariate Analysis

All variables were now included in a single model with only main effects. Variables that in this analysis, did not significantly affect a patient's chance of presenting with both active mucocutaneous and musculoskeletal disease were removed, giving the results (fitted using generalized estimating equations) shown with adjusted odds ratios and 95% confidence intervals in table 5.22.

Table 5.22: Coefficients with adjusted odds ratios from the final multivariate analysis of active musculoskeletal and mucocutaneous disease vs. inactive mucocutaneous, musculoskeletal and renal disease.

	coefficient	odds ratio
A previous occurrence of a mucocutaneous B	0.932	2.539 [1.424,4.526]
Number of musculoskeletal B scores	0.294	1.341 [1.195,1.505]
Time (years) since first clinic visit	-0.345	0.708 [0.605,0.829]

As was indicated by previous analyses a history of either mucocutaneous B scores or musculoskeletal B scores increases a patient's chance of presenting with active disease in both the mucocutaneous and musculoskeletal systems. The patient's chance increases with each additional musculoskeletal B score.

The chance of a patient presenting with active disease in both the muco-cutaneous and musculoskeletal organs/systems decreases with the time that a patient has been registered at the clinic.

Interactions

An investigation into interactions was then undertaken. No significant interactions were found.

5.1.5 Active Renal Disease with Active Disease in One or Both of the Musculoskeletal and Mucocutaneous Organs/Systems.

In this analysis a patient presenting with renal disease should be understood as a patient presenting at a clinic visit with both active renal disease and active mucocutaneous disease, or both active renal disease and active muscular disease or active disease in all three organs/systems.

Univariate Analyses

Initial univariate analyses were carried out in order to assess the relationships between history of disease activity in the organs/systems chosen for detailed analysis and a patient's chance of presenting at a clinic visit with active renal disease. The results are given with unadjusted odds ratios and p values in table 5.23.

A history of either renal B or renal A scores significantly increases a patient's chance of presenting with active renal disease.

A history of either mucocutaneous A or B scores significantly increases a patient's chance of presenting with active renal disease.

Table 5.23: Coefficients with unadjusted odds ratios and p values from the univariate analyses of active renal disease vs. inactive mucocutaneous, musculoskeletal and renal disease.

	coefficient	odds ratio	p value
A previous occurrence of a renal B	1.612	5.013	0.00003
Number of renal B scores	1.063		
(Number of renal B scores)^2	-0.127	$e^{0.936-0.254rb*}$	≈ 0
A previous occurrence of a renal B	1.836	6.271	≈ 0
Number of renal B scores	1.453		
(Number of renal B scores)^2	-0.218	$e^{1.235-0.436rb}$	≈ 0
A previous occurrence of a renal A	0.641	1.899	0.238
Number of renal A scores	0.561	1.759	0.021
A previous occurrence of a musculoskeletal B	420	1.523	0.287
Number of musculoskeletal B scores	0.109	1.115	0.117
A previous occurrence of a musculoskeletal A	0.387	1.473	0.435
Number of musculoskeletal A scores	0.387	1.473	0.071
A previous occurrence of a mucocutaneous B	0.819	2.267	0.035
Number of mucocutaneous B scores	0.232	1.261	0.005
A previous occurrence of a mucocutaneous A	0.819	2.269	0.077
Number of mucocutaneous A scores	0.465	1.592	0.007
Renal damage	1.033	2.809	0.163
Musculoskeletal damage	-0.524	0.592	0.475
Mucocutaneous damage	0.585	1.797	0.475
History of disease activity in all other organs/systems	-0.043	0.958	0.918
Time (years) since first clinic visit	0.009	1.009	0.893
Time (years) since previous clinic visit	0.472	1.603	0.085
Time (years) since last visit with active renal disease	0.028	1.028	0.860
Time (years) since last visit with active mucocutaneous disease	0.030	1.031	0.769
Time (years) since last visit with active musculoskeletal disease	-0.049	0.952	0.697

* Where rb is the number of renal Bs, that the comparison patient had been observed to score at the previous visit.

A history of disease activity in the musculoskeletal organ/system does not significantly affect a patient's chance of presenting with active renal disease. A history of disease activity in the organs/systems not chosen for detailed

analysis does not significantly affect a patient's chance of presenting with active renal disease.

None of renal, mucocutaneous or musculoskeletal damage have a significant effect on a patient's chance of presenting with active renal disease.

The times since the fist clinic visit, the previous visit, the last observation of a renal A or B, a mucocutaneous A or B, or a musculoskeletal A or B have no significant effect on a patient's chance of presenting with active renal disease.

Multivariate Analysis

Table 5.24: Coefficients with adjusted odds ratios and p values from the final multivariate analysis of active renal disease vs. inactive mucocutaneous, musculoskeletal and renal disease.

	coefficient	odds ratio
A previous occurrence of a renal B	1.730	5.643 [2.220,14.341]
Time (years) since previous clinic visit	0.597	1.817 [1.081,3.052]

All variables were now included in a single model with only main effects. Variables that in this analysis do not significantly affect a patient's chance of presenting with active renal disease were removed giving the results (fitted using generalized estimating equations) shown with adjusted odds ratios and 95% confidence intervals in table 5.24.

A history of renal B scores increases a patient's chance of presenting with active renal disease. A patient's chance of presenting with active renal disease increases with the time since the previous clinic visit.

Interactions

No significant interactions were found.

5.1.6   Goodness of fit of the Polychotomous Model.

The parameters from the simple logistic models (eqn 4.3) were then combined
to give a final polychotomous logistic model

$$P(y_{it} = k \mid y_{it-1} = k_0) = \frac{e^{\alpha_k + \beta_k x_{it}}}{\sum_{m=1}^{6} e^{\alpha_m + \beta_m x_{it}}}, \tag{5.1}$$

where k takes six distinct values corresponding to the six distinct states; in-
active disease in all organs/systems (k_0), active disease in one of the three
organs/systems only, simultaneous mucocutaneous and musculoskeletal dis-
ease, and active renal disease with either active mucocutaneous disease or
active musculoskeletal disease or both. (To identify the model β_{k_0} and α_{k_0}
are set to 0).

The goodness of fit of equation (5.1) was then assessed using the method
described in section 4.3.1. Problems occurred when in one of the 10 groups
(based on the deciles of risk) the observed number of both A and B scores
was 0. When this happened the indicator variable representing that group
was omitted from the equation and an assessment of the goodness of fit of
the model was then based on the 9 remaining groups. The results for all five
organs/systems are given in table 5.25. The results indicate that the model
appears to fit the data well.

5.2   Final Models and General Conclusions

It is of interest to note that only in the analysis of active renal disease with ac-
tive disease in one or both of the muscular or mucocutaneous organs/systems

Table 5.25: Goodness of fit test for the polychotomous model.

Active mucocutaneous disease only	10.522 (9df) (p=0.310)
Active renal disease only	2.308 (8df) (p=0.940)
Active musculoskeletal disease only	14.377 (9df) (p=0.110)
Active mucocutaneous and musculoskeletal disease	6.377 (8df) (p=0.605)
Active renal and other disease	6.532 (8df) (p=0.588)

was the time since the previous visit significant, and only in the analysis of active muscular disease was the time since the patient was last observed with active disease significant. This supports the use of the strategy employed.

The analyses presented in sections 5.1.1 to 5.1.5 do not suggest that there are any particular features in a patient's disease history that would predispose a patient to present with active disease in more than one of the three organs/systems. As a consequence of this, and, as it is of greater practical use for a doctor to have predictors of active disease in an organ/system that do not depend on the presence or absence of active disease in any other organ/system at the current visit, the analysis was repeated with only three states in which the patient might present being considered. These states being simply; active mucocutaneous disease, active musculoskeletal disease and active renal disease. No attempt was made to create a polychotomous logistic model as the outcome states are not mutually exclusive.

5.2.1 Results From Fitting the Final Models

All models have been fitted using the generalized estimating equation approach of Liang and Zeger (1986).

The confidence intervals for the odds ratios in table 5.27 were calculated using the correlation of the coefficients from the generalized linear model and

Table 5.26: Coefficients with adjusted odds ratios (with 95% CI) from the multivariate analysis of active vs. not active mucocutaneous disease.

	coefficient	odds ratio
A previous occurrence of a mucocutaneous B	0.682	1.977 [1.347,2.903]
Number of mucocutaneous B scores	0.188	1.207 [1.083,1.345]
A previous occurrence of a renal A	-0.978	0.376 [0.190,0.746]
Time (years) since first clinic visit	-0.151	0.859 [0.809,0.914]
A previous occurrence of a mucocutaneous A	1.997	-
A previous occurrence of a musculoskeletal B	0.084	-
A previous occurrence of a musculoskeletal A	-0.367	-
A previous occurrence of a renal B	-1.038	-
A previous occurrence of a mucocutaneous A* A previous occurrence of a musculoskeletal B	-1.365	-
A previous occurrence of a mucocutaneous A* A previous occurrence of a musculoskeletal A	-0.947	-
A previous occurrence of a renal B* A previous occurrence of a musculoskeletal B	1.212	-

Table 5.27: Odds ratios for combinations of variables affected by the interactions

	No previous occurrence of a mucocutaneous A	Previous occurrence of a mucocutaneous A
No history of musculoskeletal A, musculoskeletal B or renal B scores	1. 000*	7.368 [4.127,13.154]
A history of musculoskeletal A scores	0.693 [0.460,1.043]	1.981 [0.769,5.102]
A history of musculoskeletal B scores	1.088 [0.785,1.507]	2.047 [1.163,3.601]
A history of renal B scores	0.354 [0.166,0.756]	2.610 [1.348,5.055]
A history of musculoskeletal B and renal B scores	1.295 [0.789,2.125]	2.435 [1.128,4.717]
A history of musculoskeletal A and musculoskeletal B scores	0.754 [0.459,1.237]	0.550 [0.247,1.226]

* All odds rations are compared to a common default category

Table 5.28: Coefficients with adjusted odds ratios (with 95% CI) from the multivariate analysis of active vs. not active musculoskeletal disease.

	Coefficient	odds ratio
Number of musculoskeletal B scores	0.564	
(Number of musculoskeletal B scores)^2	-0.040	$e^{0.524-0.080m_ub*}$
A previous occurrence of a musculoskeletal A only	1.188	3.282 [1.600,6.733]
A previous occurrence of a renal B	-0.433	0.649 [0.471,0.894]
History of disease scores in all other organs/systems	-0.332	0.717 [0.534,0.965]
Time (years) since first clinic visit	-0.080	0.923 [0.865,0.985]
Time (years) since previous clinic visit	0.392	1.479 [1.084,2.018]
Time (years) since last visit with active musculoskeletal disease	-0.126	0.882 [0.794,0.980]
A previous occurrence of a mucocutaneous B	-0.074	-
Musculoskeletal damage	0.138	-
A previous occurrence of a musculoskeletal A	0.003	-
A previous occurrence of a mucocutaneous A	-0.672	-
A previous occurrence of a mucocutaneous B* Musculoskeletal damage	0.605	-
A previous occurrence of a musculoskeletal A* A previous occurrence of a mucocutaneous A	1.157	-

* Where m_ub is the number of musculoskeletal Bs, that the comparison patient had been observed to score at the previous visit.

	No musculoskeletal damage	Musculoskeletal damage
No previous occurrence of a mucocutaneous B score	1	1.148 [0.706,1.866]
Previous occurrence of a mucocutaneous B score	0.929 [0.674,1.280]	1.951 [1.174,3.243]

	No Previous occurrence of a musculoskeletal A score	Previous occurrence of a musculoskeletal A score
No previous occurrence of a mucocutaneous A score	1	1.003* [0.633,1.590]
Previous occurrence of a mucocutaneous A score	0.511 [0.289,0.901]	1.630 [1.050,2.528]

*this odds ratio should be interpreted with caution as the effects of musculoskeletal A and B scores have already been adjusted for

the robust variance-covariance matrix from the model fitted using generalized estimating equations.

Table 5.29: Coefficients with adjusted odds ratios (with 95% CI) from the final multivariate analysis of active vs. not active renal disease.

	coefficient	odds ratio
A previous occurrence of a renal B	2.243	9.426 [4.274,20.789]
A previous occurrence of a renal A only	3.110	22.420 [8.342,60.253]
Number of renal B scores	0.254	1.289 [1.100,1.512]
Renal damage	1.222	3.393 [2.117,5.438]
A previous occurrence of a musculoskeletal B	-0.541	0.582 [0.392,0.865]
Time (years) since previous clinic visit	0.496	1.642 [1.085,2.485]
Time (years) since last visit with active renal disease	-0.315	0.730 [0.569,0.936]

5.2.2   Goodness of Fit of the Final Models

Table 5.30: Goodness of fit tests for individual logistic models.

Active mucocutaneous disease	6.920 (9df) (p=0.645)
Active renal disease	7.388 (9df) (p=0.597)
Active musculoskeletal disease	8.514 (9df) (p=0.483)

All three final models appear to fit the data well. There are only minor differences between the results from these individual logistic regression analyses and the results from the polychotomous model. These were that, in the simple logistic model for active renal disease, time since the last visit and time since the previous renal A or B score are significant and were not in

the polychotomous model. Also in the simple logistic model for active renal disease a history of musculoskeletal B scores is adequately represented by a previous occurrence of a musculoskeletal B as opposed to the number of musculoskeletal B's. In the simple logistic model for active musculoskeletal disease time since the first clinic visit and time since the last visit are significant and were not in the polychotomous model. In the simple logistic model for active mucocutaneous disease there is a significant interaction between musculoskeletal damage and a previous occurrence of a mucocutaneous B that was not present in the polychotomous model.

There does appear to be greater evidence of dependence on the time between visits in the simple logistic models. However the effects are not large and the time between visits has been adjusted for by the inclusion of $t'_{ip} - (t'_{ip-1})$ as discussed in section 4.2.1. There does also appear to be evidence in the simple logistic models of dependence on the time since the patients were last observed with active disease. However again the effects are small.

5.2.3 Assessing the Assumption of Independence of Activity in the Three Organ/Systems

If, at any point in time, the development of activity in any one of the three organs/systems were to be regarded as independent (conditional on past activity) of the development of disease activity in any of the others, then a 6 category polychotomous model could be calculated from the three simple logistic models. In order to assess the validity of this assumption, the estimates of risk for each observation were calculated using the simple logistic models. Estimates of risk were then calculated from these (based on the assumption of independence), for active mucocutaneous disease only, active

129

musculoskeletal disease only, active renal disease only, active mucocutaneous and active musculoskeletal disease and finally active renal disease with active disease in one or both of the other organs/systems. The goodness of fit of the polychotomous model was then tested using the method described in section 4.3.1 where a subject's group membership was based on these estimates of risk. The results are given in table 5.31.

Table 5.31: Goodness of fit test for polychotomous logistic model calculated form the three simple logistic models.

Active mucocutaneous disease only	9.409 (9df) (p=0.400)
Active renal disease only	14.654 (9df) (p=0.101)
Active musculoskeletal disease only	14.946 (9df) (p=0.092)
Active mucocutaneous and musculoskeletal disease	15.010 (9df) (p=0.091)
Active renal and other disease	13.653 (6df) (p=0.034)

The results given in table 5.31 generally support the assumption that disease activity in any of the three organs/systems is independent (conditional on past activity) of disease activity in any of the others. However it should perhaps be noted that although activity in the mucocutaneous organ/system and activity in the musculoskeletal organ/system do appear to be independent (conditional on past activity) there is perhaps some evidence that activity in the renal organ/system is not independent (conditional on past activity) of activity in either the mucocutaneous or musculoskeletal organs/systems. The total number of observations in this category is however small (table 4.3).

5.2.4 Medication

As discussed in section 4.3, as a final stage in the variable selection, medication was added to the models, to assess whether any of the results found changed once the treatment a patient was receiving had been accounted for. Variables were created that indicated what medication a patient was receiving at the previous clinic visit. For this purpose four groups of treatments were considered. NSAIDs, steroids, immunosuppressants and other drugs.

As this is not a randomized controlled trial no discussion will be made of any apparent effects of the medication, and this final stage in the model selection is, as stated, simply to assess whether any of the observed effects are influenced by treatments patients receive.

Tables 5.32, 5.33 and 5.34 show the coefficients and significance levels for the three final models with the medication a patient was receiving at the previous visit having been included.

The results presented in tables 5.32, 5.33 and 5.34 indicated that the inclusion of a patient's treatment in the models did not affect the results from 5.2.1.

It should be remembered that all the models are based on a subset of visits where the patient was known to have inactive disease at the previous visit and that consequently the effects of the medications should be interpreted as the effect of the medication on preventing active disease.

The apparent lack of effect of the treatment may indicate that the physicians have made the correct decision about the treatment received.

It is of interest to note that a similar study in psoriatic arthritis (Gladman *et al.* 1995) also found that the inclusion of medication in models for severe disease did not alter the apparent effects of other variables.

Table 5.32: Coefficients with significance levels from the final multivariate analysis of active vs. not active renal disease with medication at the previous visit.

	coefficient	significance level
A previous occurrence of a renal B	2.063	≈ 0
A previous occurrence of a renal A only	2.899	≈ 0
Number of renal B scores	0.240	0.003
Renal damage	1.177	≈ 0
A previous occurrence of a musculoskeletal B	-0.465	0.025
Time (years) since previous clinic visit	0.537	0.015
Time (years) since last visit with active renal disease	-0.308	0.017
NSAIDS	-0.263	0.268
Steroids	0.124	0.661
Immunosuppressants	0.404	0.045
Other medication	-3.00	≈ 0

5.3 Validating the Models

A second data set was used to validate the three final logistic models. This data set is from the Department of Rheumatology at the Middlesex Hospital and consists of demographic information, the BILAG activity index, the SLICC damage index and information on therapy and laboratory data.

The data has been collected on 295 patients, 265 women and 30 men over a period of five years. 190 of the patients are Caucasian, 52 Afro-Caribbean, 32 Asian and 21 of other ethnic origins.

The number of visits involved are given in table 5.35.

Table 5.33: Coefficients with significance levels from the multivariate analysis of active vs. not active mucocutaneous disease with medication at the previous visit.

	coefficient	significance level
A previous occurrence of a mucocutaneous B	0.699	0.0003
Number of mucocutaneous B scores	0.196	0.0004
A previous occurrence of a renal A	-0.963	0.006
Time (years) since first clinic visit	-0.158	≈ 0
A previous occurrence of a mucocutaneous A	2.032	≈ 0
A previous occurrence of a musculoskeletal B	0.104	0.535
A previous occurrence of a musculoskeletal A	-0.330	0.132
A previous occurrence of a renal B	-1.003	0.013
A previous occurrence of a mucocutaneous A*		
A previous occurrence of a musculoskeletal B	-1.397	0.001
A previous occurrence of a mucocutaneous A*		
A previous occurrence of a musculoskeletal A	-0.897	0.034
A previous occurrence of a renal B*		
A previous occurrence of a musculoskeletal B	1.190	0.007
NSAIDS	0.100	0.480
Steroids	-0.219	0.142
Immunosuppressants	0.062	0.683
Other medication	-5.133	≈ 0

Table 5.34: Coefficients with significance levels from the multivariate analysis of active vs. not active musculoskeletal disease with medication at the previous visit.

	Coefficient	significance level
Number of musculoskeletal B scores	0.576	
(Number of musculoskeletal B scores)^2	-0.041	≈ 0
A previous occurrence of a musculoskeletal A only	1.208	0.0008
A previous occurrence of a renal B	-0.447	0.008
History of disease scores in all other organs/systems	-0.305	0.051
Time (years) since first clinic visit	-0.084	0.0001
Time (years) since previous clinic visit	0.379	0.021
Time (years) since last visit with active musculoskeletal disease	-0.124	0.022
A previous occurrence of a mucocutaneous B	-0.071	0.661
Musculoskeletal damage	0.134	0.589
A previous occurrence of a musculoskeletal A	0.011	0.961
A previous occurrence of a mucocutaneous A	-0.686	0.018
A previous occurrence of a mucocutaneous B*		
Musculoskeletal damage	0.583	0.047
A previous occurrence of a musculoskeletal A*		
A previous occurrence of a mucocutaneous A	1.173	0.002
NSAIDS	0.041	0.754
Steroids	-0.269	0.087
Immunosuppressants	0.204	0.151
Other medication*		

* No patients with inactive disease in all systems receiving 'other' medication were observed at the next clinic visit with active muscular disease.

Table 5.35: Number of visits (in the second data set) where the patient presented with active disease in one or more organs/systems having had inactive disease in all systems at their previous visit.

Inactive renal mucocutaneous and musculoskeletal disease	$\frac{1892}{2341} = 80.82\%$
Active mucocutaneous disease only	$\frac{158}{2341} = 6.75\%$
Active musculoskeletal disease only	$\frac{148}{2341} = 6.32\%$
Active renal disease only	$\frac{95}{2341} = 4.38\%$
Active mucocutaneous and musculoskeletal disease	$\frac{25}{2341} = 2.26\%$
Active renal disease with either active mucocutaneous, musculoskeletal or both	$\frac{23}{2341} = 0.98\%$

5.3.1 The three final models fit using the second data set

The models were fit using the second data set; the results are given in tables 5.37, 5.38 and 5.39. Only the coefficients are presented as these results are included in the text simply for the purpose of comparison.

It was not possible to include the interaction between a previous occurrence of a mucocutaneous A and a previous occurrence of a musculoskeletal A in the analysis of active vs. not active mucocutaneous disease as no patients in this data set were observed with both a history of mucocutaneous A scores and a history of musculoskeletal A scores.

Table 5.36: Goodness of fit tests for individual logistic models fit using the second data set.

Active mucocutaneous disease	16.600 (9df) (p=0.103)
Active renal disease	11.478 (9df) (p=0.244)
Active musculoskeletal disease	14.194 (9df) (p=0.116)

Table 5.37: Coefficients from the multivariate analysis of active vs. not active mucocutaneous disease using the second dataset.

	coefficient	standard error
A previous occurrence of a mucocutaneous B	0.239	0.267
Number of mucocutaneous B scores	0.468	0.130
A previous occurrence of a renal A	0.383	0.679
Time (years) since first clinic visit	0.0002	0.00009
A previous occurrence of a mucocutaneous A	2.569	0.593
A previous occurrence of a musculoskeletal B	0.063	0.223
A previous occurrence of a renal B	-0.463	0.259
A previous occurrence of a mucocutaneous A* A previous occurrence of a musculoskeletal B A previous occurrence of a renal B*	-2.335	0.691
A previous occurrence of a musculoskeletal B	-0.268	0.403

Table 5.38: Coefficients from the multivariate analysis of active vs. not active renal disease using the second data set.

	coefficient	standard error
A previous occurrence of a renal B	1.411	0.478
Number of renal B scores	0.038	0.160
Renal damage	0.695	0.317
History of musculoskeletal B scores	-0.221	0.149
Time (years) since previous clinic visit	0.384	0.200
Time (years) since last visit with active renal disease	-0.720	0.253

It was also not possible to include a variable for renal A scores only in the analysis of active vs. not active renal disease as no patients in this data set scored a renal A and not a renal B.

Table 5.39: Coefficients from the multivariate analysis of active vs. not active musculoskeletal disease using the second dataset.

	Coefficient	standard error
Number of musculoskeletal B scores	0.989	0.264
(Number of musculoskeletal B scores)^2	-0.171	0.061
A previous occurrence of a musculoskeletal A only	1.209	0.562
A previous occurrence of a renal B	-0.449	0.249
History of disease scores in all other organs/systems	0.162	0.221
Time (years) since first clinic visit	-0.224	0.079
Time (years) since previous clinic visit	0.505	0.188
Time (years) since last visit with active musculoskeletal disease	0.189	0.146
A previous occurrence of a mucocutaneous B	0.296	0.244
Musculoskeletal damage	0.625	0.272
A previous occurrence of a musculoskeletal A	1.117	0.315
A previous occurrence of a mucocutaneous A	-0.154	0.383
A previous occurrence of a mucocutaneous B* Musculoskeletal damage	-0.348	0.461
A previous occurrence of a musculoskeletal A* A previous occurrence of a mucocutaneous A	0.361	0.553

* Where $m_u b$ is the number of musculoskeletal Bs, that the comparison patient had been observed to score at the previous visit.

The results of the goodness of fit test applied to the models fit using this second data set are given in table 5.36.

In order to assess whether there were significant differences between the two data sets, the models were fit using a combined data set and the inter-

actions between hospital and all the covariates were examined. The results of global tests of the interactions are given in table 5.40.

In the mucocutaneous model the significant differences are that, the number of mucocutaneous B's the patient has scored has a larger effect on a patient's chance of presenting with active mucocutaneous disease if the patient is from the Department of Rheumatology at the Middlesex Hospital, and, patients from this department do not have an increased chance of presenting with active mucocutaneous disease if they have a history of both active renal and musculoskeletal disease.

In the musculoskeletal model the significant differences are that time, time since the previous visit and time since the patient last presented with active musculoskeletal disease have different effects on the patients from the Department of Rheumatology at the Middlesex Hospital. The effect of the interaction between a previous occurrence of a musculoskeletal A and previous occurrence of a mucocutaneous A is also different. Patients from the Birmingham hospitals have an decreased chance of of presenting with active musculoskeletal disease if they have a history of mucocutaneous A scores, unless they also have a history of musculoskeletal A scores in which case their chance of presenting with active disease increases. Patients from the Department of Rheumatology at the Middlesex Hospital have an increased chance of presenting with musculoskeletal disease if they have a history of mucocutaneous A scores irrespective of whether they have a history of musculoskeletal A scores.

In the renal model all the effects seem to be smaller if the patient is from the Department of Rheumatology at the Middlesex Hospital, with the number of renal B scores having no effect on a patient's chance of presenting with active renal disease.

Table 5.40: Tests of interaction between hospital and all covariates in the models.

Active mucocutaneous disease	32.028 (9df) (p=0.0002)
Active renal disease	19.176 (6df) (p=0.004)
Active musculoskeletal disease	48.273 (14df) (p=0.00001)

Nevertheless the majority of the findings from the first data set are qualitatively supported by the analysis of this second data set. Some of the differences that do exist may simply reflect differences in patient management.

5.4 Summary of the results found

In interpreting the results of the analyses, it is important to remember that all patients with a history of mucocutaneous A scores also have a history of mucocutaneous B scores, and very few patients have only histories of renal A and musculoskeletal A scores (tables 5.10, 5.17). Consequently when a history of B scores is represented in a model, and a history of A scores is not, this simply means that the size of the effect does not increase with severity of disease, not, that a history of A scores does not have an effect. Where patients have only a history of A scores (not B scores) the numbers are generally too small for any significance to be achieved when included in a model with a history of B scores. Recall that further details of variable selection, and discussion of why some effects disappear, can be found in appendix D.

As a consequence of this, for the purpose of the subsequent discussion, a history of B scores has been interpreted as a history of active disease. Where

a history of A scores has a specific effect, not evident in those patients with only a history of B scores, this should be interpreted as only severe disease activity having the effect.

This analysis shows that patients with a history of disease activity in any of the mucocutaneous, musculoskeletal or renal organs/systems are more likely to continue to have active disease in that organ/system than patients with no previous evidence of disease activity (in that organ/system). In addition patients with a greater number of clinic visits with active disease are generally more likely to develop active disease than those with less.

This analysis also supports the hypothesis that a possible subset of lupus is represented by patients with renal disease and little else. A history of renal activity appears to decrease patients' chances of developing both active mucocutaneous and musculoskeletal disease. This analysis also suggests that a patient is less likely to present with renal activity if they have a history of active musculoskeletal disease. A patient's chance of presenting with renal activity is also increased if the patient has renal damage. This particular finding is possibly due to the fact that renal activity is often 'clinically silent'. That is to say that evidence for renal activity is, in the early phases, only indicated by the presence of protein or small amounts of blood in the urine which are only detectable on dipstick testing. Likewise an increase in blood pressure does not cause symptoms in the early stages. The patient may therefore enter the stage of permanent damage before clinical features become evident. From this analysis, it is therefore apparent that renal disease is most likely to occur on its own, with no evidence that disease activity or damage in any other system increases the chance of active renal disease.

There also appear to be a number of associations between activity in the musculoskeletal and mucocutaneous systems, supporting the suggestion that

a possible subset of lupus is represented by patients with musculoskeletal and mucocutaneous disease alone. The analysis suggests that if a patient has a history of mucocutaneous activity and musculoskeletal damage their chance of presenting with active musculoskeletal disease is increased, and that patients with a history of severe mucocutaneous and musculoskeletal disease have an increased chance of developing active musculoskeletal disease.

Further associations between activity in these two systems were also identified. It appears that patients with a history of severe (BILAG A scores) mucocutaneous activity are less likely to develop subsequent active mucocutaneous disease if they have a history of musculoskeletal activity. It also appears that patients who have previously scored a BILAG A in the mucocutaneous system are less likely to develop active musculoskeletal disease as long as they have not previously scored a BILAG A in the musculoskeletal system.

The length of time that a patient has been registered at a clinic is associated with a reduction in both mucocutaneous and musculoskeletal activity. However a patient's chance of developing active musculoskeletal or renal disease increases with the time since the last visit but decreases with time since the patient was last observed to score an A or B in the respective system.

Chapter 6

Efficiency

The lack of available information on the times of failure, and the need to investigate interrelationships between different systems led to the use of logistic models for the analysis of the lupus data. Dynamic covariates were used without any appeal to an underlying model in continuous time. As discussed in section 4.2 the use of these discrete models as opposed to the 'true' model, will undoubtedly have led to some loss of information. In this chapter, consideration is given to the possible extent of this loss of information.

The approach taken is to assume a model for the underlying disease process in continuous time. If this model was, in fact, known then it would typically be used for inference and would make full use of the information available in the data. Given this model, assumed to be the 'true' model, the efficiency of the use of an incorrect model for the testing of specific hypotheses can be examined. A relatively simple form for the correct model is assumed in order to permit analytical results to be derived but the qualitative conclusions should be more widely applicable.

Assume therefore that the times between active disease in each individual organ/system are exponentially distributed and that all patients begin follow-

up with no active disease. Further assume that there are two groups of patients, the first of size n_1 and the second of size n_2 (n_1 and n_2 are fixed with $n_1 + n_2 = n$). Let the patients in the second group have a characteristic that distinguishes them from those in the first. $T_1, ..., T_n$ can then be defined as independent random variables representing the time until the occurrence of active disease (in one organ/system) for each of the n patients. The 'true' model for active disease in an organ/system is characterised by the probability density functions for the T's which, in this situation, take the form

$$f(t|\lambda_i) = \lambda_i e^{-\lambda_i t} \quad t \geq 0 \tag{6.1}$$

for $i = 1, 2$, and where $\lambda_1 = \lambda$ is the rate parameter for the patients in the first group and $\lambda_2 = \lambda e^\gamma$ is the rate parameter for the patients in the second group. The data available for analysis will therefore relate to two independent samples, with possibly different means, of exponential random variables. These times will, however, not be observed directly and the information available on the times will depend on the pattern of observation of the panel data.

In chapter 5 the logistic models used were of the form

$$\text{logit} \Pr(y_{pt'} = 1) = \alpha + \beta x_{pt'},$$

where $\text{logit} \Pr(y_{pt'} = 1) = \log \frac{\Pr(y_{pt'}=1)}{1-\Pr(y_{pt'}=1)}$, and $y_{pt'}$ is the response (active disease or not) at time t' of the pth of K subjects. (In some models the vector $x_{pt'}$ contained the actual gap time between visits. That situation will not be considered in any detail in this chapter). When, as in the situation considered here, the patients fall into two distinct groups, and the effect of

143

only group membership is being considered, the equivalent logistic model would take the form

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \alpha + \beta x_i \tag{6.2}$$

for i = 1,2 where

$$\pi_i = \frac{e^{(\alpha+\beta x_i)}}{1 + e^{(\alpha+\beta x_i)}}$$

is the probability of a patient in group i presenting with active disease at their next clinic visit, or equivalently, developing active disease between two clinic visits. In equation (6.2) x_i is a binary variable indicating whether the patient has the characteristic that distinguishes the patients in the second group, or not. That is $x_i = 0$ for patients in the first group and $x_i = 1$ for patients in the second group.

It is natural to focus attention on the test of the null hypothesis of no group differences, that is represented by the null hypotheses $\hat{\gamma} = 0$ and $\hat{\beta} = 0$ in the exponential and logistic models respectively. Since, in general, the parameters γ and β do not estimate the same quantity, an asymptotic comparison of the tests of the null hypothesis of no group differences in the two models is most naturally done through calculation of Pitman efficiency. The use of efficiency in this context is recommended by Cox (1957). Pitman efficiency for example, has been used by Farewell (1982) to study the effect of ignoring variability in the misclassification of an ordinal response, by Matthews (1984) to study the use of an incorrect model for competing risks, by Lagakos et al. (1978) to study the effect of mismodelling an explanatory variable in regression models and by Farewell et al. (2004) to study interaction tests in generalized linear models. While Pitman efficiency is an asymptotic measure, it should provide a qualitative indication of the relative behaviour of the two models.

144

Pitman Efficiency

Consider two consistent asymptotically normal statistics S_1 and S_2 with limiting variances $n_1^{-1}\sigma_{S_1}^2(\gamma_0)$ and $n_2^{-1}\sigma_{S_2}^2(\gamma_0)$. S_1 and S_2 have asymptotically the same power for the same local values of γ if the sample sizes n_1 and n_2 are related by

$$n_1^{-1}\sigma_{S_1}^2(\gamma_0) = n_2^{-1}\sigma_{S_2}^2(\gamma_0).$$

Thus if efficiency is measured by relative sample sizes, the asymptotic relative efficiency of S_1 to S_2, with respect to the test of the null hypothesis $\gamma = \gamma_0$, is

$$e(S_1 : S_2) = \frac{\sigma_{S_2}^2(\gamma_0)}{\sigma_{S_1}^2(\gamma_0)}.$$

However it is unnecessary to restrict S_1 and S_2 to be consistent estimates of γ. More generally it might be assumed that S_i ($i = 1, 2$) are consistent for some function $\mu_{S_i}(\gamma_0)$ which is a monotone function of γ_0, at least near to $\gamma = \gamma_0$. In this situation, Pitman efficiency provides the corresponding general definition of asymptotic relative efficiency (Cox and Hinkley 1974). Pitman efficiency can be used for statistics S_i that are asymptotically $N(\mu_{S_i}(\gamma), \sigma_{S_i}^2(\gamma))$. The asymptotic relative efficiency of two such statistics S_1 and S_2 is given by

$$e(S_1 : S_2) = \left(\frac{\mu_{S_1}'(\gamma_0)}{\mu_{S_2}'(\gamma_0)}\right)^2 \left(\frac{\sigma_{S_2}^2(\gamma_0)}{\sigma_{S_1}^2(\gamma_0)}\right). \tag{6.3}$$

with $\mu_{S_i}'(\gamma_0)$ denoting the derivative of the function $\mu_{S_i}(\gamma_0)$.

In the situation considered here, use of the exponential model will provide a consistent estimate of γ. It is then possible to calculate the relative efficiency of the logistic model to the exponential from equation (6.3) based

on a statistic from the logistic model analysis, S_1, that is asymptotically distributed as $N(\beta(\gamma), \frac{V(\beta(\gamma))}{n})$ and a statistic from the exponential model, S_2, that is asymptotically distributed as $N(\gamma, \frac{V(\gamma)}{n})$.

Recall that, in general, the interpretation of γ is not the same as the interpretation of β and it would not be reasonable to attempt to conclude anything from equation (6.3) for any values of γ and β other than zero. While there is a formal link between β and γ for all values of γ, knowledge of the link requires knowledge of the 'true' model and that is what, in practice, is unknown.

The relative efficiency of the logistic model will be affected by the number of different times at which patients are observed. A number of different situations are looked at in sections 6.1 to 6.4.

6.1   One Time Point

If all the patients from each group are observed at the same time then no efficiency should be lost by the use of a logistic model as opposed to an exponential model. As the number of time points that the patients are assessed at increases, the relative efficiency of the logistic models to the exponential models would be expected to decrease. In order to set up the structure of the approach the simplest case was initially examined where it was assumed that all patients in both groups were observed at a single fixed time point t. Let z_1 represent the number observed to have failed (that is to have developed active disease) in the first group and z_2 in the second.

6.1.1   The Exponential Model

The log-likelihood l^* for the exponential model (equation (6.1)) is

$$l^* = \sum_{i=1}^{2}[(n_i - z_i)(-\lambda_i t) + z_i \log(1 - e^{-\lambda_i t})], \qquad (6.4)$$

for i=1,2 and with $\lambda_1 = \lambda$. This gives the variances of $\hat\lambda$ and $\hat\lambda_2$ as

$$V(\hat\lambda) = \frac{1 - e^{-\hat\lambda t}}{n_1 t^2 e^{-\hat\lambda t}}$$

$$V(\hat\lambda_2) = \frac{1 - e^{-\hat\lambda_2 t}}{n_2 t^2 e^{-\hat\lambda_2 t}}.$$

$$V(\hat\gamma) = V(\log(\frac{\hat\lambda}{\hat\lambda_2}))$$

$$V(\log(\frac{\hat\lambda}{\hat\lambda_2})) = V(\log(\hat\lambda) - \log(\hat\lambda_2))$$

$$V(\log(\hat\lambda) - \log(\hat\lambda_2)) = V(\log(\hat\lambda)) + V(\log(\hat\lambda_2)).$$

Given that we can approximate the variance of a function $g(\mathbf{X})$ by

$$Var_\vartheta g(\mathbf{X}) \approx E_\vartheta([g(\mathbf{X}) - g(\vartheta)]^2)$$

$$\approx E_\vartheta((\sum_{\delta=1}^{k} g_\delta'(\vartheta)(X_\delta - \vartheta_\delta))^2)$$

$$= \sum_{\delta=1}^{k}[g_\delta'(\vartheta)]^2 Var_\vartheta X_\delta + 2\sum_{\delta > \epsilon} g_\delta'(\vartheta)g_\epsilon'(\vartheta)Cov_\vartheta(X_\delta, X_\epsilon)$$

we get

$$V(\hat\gamma) \approx \frac{1}{\hat\lambda^2}V(\hat\lambda) + \frac{1}{\hat\lambda_2^2}V(\hat\lambda_2). \qquad (6.5)$$

Giving

$$V(\hat\gamma) \approx \frac{1}{\hat\lambda_2^2 t^2}[\frac{n_1(1 - e^{-\hat\lambda_2 t})e^{-\hat\lambda t} + n_2(1 - e^{-\hat\lambda t})e^{-\hat\lambda_2 t}e^{2\hat\gamma}}{n_1 n_2 e^{-\hat\lambda t}e^{-\hat\lambda_2 t}}]. \qquad (6.6)$$

6.1.2 The Logistic Model

As previously discussed the characteristic of interest that distinguishes the patients in the second group from those in the first, is represented in the logistic model (equation (6.2)) by a binary indicator variable x_i, with corresponding parameter β; $x_1 = 0$ for those patients in group 1, and $x_2 = 1$ for those patients in group 2.

The log-likelihood is

$$l = \sum_{i=1}^{2} [z_i(\alpha + \beta x_i) - n_i \log(1 + e^{(\alpha + \beta x_i)})] \qquad (6.7)$$

(where z_i is the number of patients in group i observed to fail (that is to say have active disease) at time t.) The likelihood equations for estimating

$$\Omega = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

are (for j=1,2)

$$\frac{\partial l(\hat{\Omega})}{\partial \Omega_j} = 0, \qquad (6.8)$$

That is

$$\frac{\partial l(\hat{\Omega})}{\partial \alpha} = 0,$$

and

$$\frac{\partial l(\hat{\Omega})}{\partial \beta} = 0.$$

Assume that for a given

$$\Lambda = \begin{pmatrix} \lambda \\ \gamma \end{pmatrix},$$

$$\hat{\Omega} = \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix}$$

148

converges in probability as $n \rightarrow \infty$ to a limit,

$$\Omega_\Lambda = \begin{pmatrix} \alpha_\Lambda \\ \beta_\Lambda \end{pmatrix},$$

and let for notational convenience (for j,k=1,2)

$$G_j = \frac{\partial l(\Omega_\Lambda)}{\partial \Omega_j},$$

and

$$G_{jk} = \frac{\partial l(\Omega_\Lambda)}{\partial \Omega_j \partial \Omega_k}.$$

From (Cox 1961) expanding equation (6.8) asymptotically we have that

$$\frac{1}{n} G_j + \sum_{k=1}^{2} \left[E_\Lambda(G_{jk})(\hat{\Omega}_k - \Omega_{\Lambda,k}) \right] = 0, \qquad (6.9)$$

where E_Λ denotes the expectation under the underlying distribution (the exponential). Since $\hat{\Omega}$ converges to Ω_Λ it follows that

$$E_\Lambda(G_j) = 0. \qquad (6.10)$$

It follows from equation (6.9) that

$$\hat{\Omega}_k - \Omega_{\Lambda,k} = -\frac{1}{n} \sum_{j=1}^{2} (E_\Lambda(G^{..}))^{jk} G_j, \qquad (6.11)$$

where $(E_\Lambda(G^{..}))^{jk}$ is the j, k^{th} element of the matrix inverse to $E_\Lambda(G_{jk})$. Using the standard result that

$$E_\Lambda(G_j G_k) + E_\Lambda(G_{jk}) = 0$$

it follows from equation (6.11) that

$$Cov(\hat{\Omega}_j - \Omega_{\Lambda,j}, \hat{\Omega}_k - \Omega_{\Lambda,k})$$
$$= \frac{1}{n} \sum_{l=1}^{2} \sum_{m=1}^{2} (E_\Lambda(G^{..}))^{jl} E_\Lambda(G_l G_m)(E_\Lambda(G^{..}))^{mk}. \qquad (6.12)$$

149

For the logistic model under consideration, from equation (6.10), letting $j = 1$, we get

$$E_{\Lambda}(\sum_{i=1}^{2}(z_i - n_i\pi_i)) = 0,$$

and, letting $j = 2$,

$$E_{\Lambda}(\sum_{i=1}^{2}x_i(z_i - n_i\pi_i)) = 0,$$

where,

$$E_{\Lambda}(z_1) = n_1(1 - e^{-\lambda t})$$

and

$$E_{\Lambda}(z_2) = n_2(1 - e^{-\lambda_2 t}).$$

This gives

$$\hat{\alpha}_{\Lambda} = \log[e^{\lambda t} - 1]$$

and

$$\hat{\beta}_{\Lambda} = \log[\frac{e^{\lambda_2 t} - 1}{e^{\lambda t} - 1}]. \tag{6.13}$$

From equation (6.12) we get

$$V_{\Lambda}(\hat{\beta}) = \frac{1}{n_1(1 - e^{-\lambda t})e^{-\lambda t}} + \frac{1}{n_2(1 - e^{-\lambda_2 t})e^{-\lambda_2 t}}. \tag{6.14}$$

6.1.3 The Relative Efficiency of the Logistic Model

From equations (6.6), (6.13) and (6.14), and substitution into equation (6.3) with S_1 asymptotically $N(\beta_{\Lambda}, \frac{V_{\Lambda}(\beta)}{n})$ and S_2 asymptotically $N(\gamma, \frac{V(\gamma)}{n})$ the Pitman relative efficiency of the two models (the logistic and the exponential) is

$$e(S_1 : S_2) = \frac{(1 - e^{-\lambda t})[n_1(1 - e^{-\lambda_2 t})e^{-\lambda t} + n_2 e^{2\gamma}(1 - e^{-\lambda t})e^{-\lambda_2 t}]}{(1 - e^{-\lambda_2 t})[n_1(1 - e^{-\lambda t})e^{-\lambda t} + n_2(1 - e^{-\lambda_2 t})e^{-\lambda_2 t}]}. \tag{6.15}$$

Equation (6.15) is only of interest when $\gamma = 0$, in which case $e(S_1 : S_2) = 1$. This shows that, as expected, when the patients are all observed at the same time the logistic model is as efficient as the exponential in testing the null hypothesis of no effect.

6.2 Two Time Points

As previously discussed the relative efficiency of the logistic model to the exponential will depend on the number of different time points at which the patients in each group are assessed. To initially examine this, it was assumed that half the patients in each group were observed at time t_1 and half at time t_2 (note that the notation here is different from the notation used in 4.2.1). To simplify the calculations each group was taken to have the same number of patients n' ($n' = \frac{n}{2}$). In this situation the logistic model would commonly be stratified by time, and the model used in section 6.2.2 is naive. However this model is used to illustrate the effect of an increasing number of time points on the relative efficiency of the models used in chapter 5 when initially the logistic models do not incorporate time.

6.2.1 The Exponential Model

Note that while we assume that the exact failure times derive from an exponential model we still only have panel data. Thus the likelihood incorporates expressions for the probability of a patient in group i failing before time t_j, which for the exponential model are

$$\int_0^{t_j} \lambda_i e^{-\lambda_i t} dt = 1 - e^{-\lambda_i t_j}$$

$$i = 1, 2$$

$$j = 1, 2.$$

With z_{ij} the number of patients observed to fail in group i before time t_j, the log-likelihood l^* is

$$l^* = \sum_{j=1}^{2} \sum_{i=1}^{2} [(\frac{n'}{2} - z_{ij})(-\lambda_i t_j) + z_{ij} \log(1 - e^{-\lambda_i t_j})].$$

With $\lambda_1 = \lambda$ this gives

$$v(\hat{\lambda}) = \frac{2(1 - e^{-\hat{\lambda} t_1})(1 - e^{-\hat{\lambda} t_2})}{n'[t_1^2(1 - e^{-\hat{\lambda} t_2})e^{-\hat{\lambda} t_1} + t_2^2(1 - e^{-\hat{\lambda} t_1})e^{-\hat{\lambda} t_2}]}$$

$$v(\hat{\lambda}_2) = \frac{2(1 - e^{-\hat{\lambda}_2 t_1})(1 - e^{-\hat{\lambda}_2 t_2})}{n'[t_1^2(1 - e^{-\hat{\lambda}_2 t_2})e^{-\hat{\lambda}_2 t_1} + t_2^2(1 - e^{-\hat{\lambda}_2 t_1})e^{-\hat{\lambda}_2 t_2}]},$$

which, using equation (6.5), leads to

$$v(\hat{\gamma}) = \frac{1}{\hat{\lambda}^2}[\frac{2(1 - e^{-\hat{\lambda} t_1})(1 - e^{-\hat{\lambda} t_2})}{n'[t_1^2(1 - e^{-\hat{\lambda} t_2})e^{-\hat{\lambda} t_1}) + t_2^2(1 - e^{-\hat{\lambda} t_1})e^{-\hat{\lambda} t_2}]}] +$$

$$\frac{1}{\hat{\lambda}_2^2}[\frac{2(1 - e^{-\hat{\lambda}_2 t_1})(1 - e^{-\hat{\lambda}_2 t_2})}{n'[t_1^2(1 - e^{-\hat{\lambda}_2 t_2})e^{-\hat{\lambda}_2 t_1} + t_2^2(1 - e^{-\hat{\lambda}_2 t_1})e^{-\hat{\lambda}_2 t_2}]}]. \qquad (6.16)$$

6.2.2 The Logistic Model

We now consider the use of a logistic model for panel data. From the logistic model the probability of a patient in group i failing before time t_j would be written as

$$\pi_i = \frac{e^{(\alpha + \beta x_i)}}{1 + e^{(\alpha + \beta x_i)}}.$$

With again z_{ij} being the number of patients observed to fail in group i before time t_j, the log-likelihood is

$$l = \sum_{j=1}^{2} \sum_{i=1}^{2} [z_{ij}(\alpha + \beta x_i) - \frac{n'}{2} \log(1 + e^{(\alpha + \beta x_i)})]. \qquad (6.17)$$

As previously the characteristic of interest is represented in the logistic model by a binary indicator variable x_i, with corresponding parameter β; $x_1 = 0$ for

those patients in group 1, and $x_2 = 1$ for those patients in group 2. (Equation (6.17) is equivalent to equation (6.7) and, as stated previously, this analysis could possibly be considered naive). In this case, equation (6.10) with

$$E_\Lambda(z_1) = \frac{n'}{2}(1 - e^{-\lambda t_1}) + \frac{n'}{2}(1 - e^{-\lambda t_2})$$

and

$$E_\Lambda(z_2) = \frac{n'}{2}(1 - e^{-\lambda_2 t_1}) + \frac{n'}{2}(1 - e^{-\lambda_2 t_2})$$

(where $z_1 = z_{11} + z_{12}$ and $z_2 = z_{21} + z_{22}$) gives,

$$\hat{\alpha}_\Lambda = \log[\frac{2 - e^{-\lambda t_1} - e^{-\lambda t_2}}{e^{-\lambda t_1} + e^{-\lambda t_2}}]$$

and

$$\hat{\beta}_\Lambda = \log[\frac{(2 - e^{-\lambda_2 t_1} - e^{-\lambda_2 t_2})(e^{-\lambda t_1} + e^{-\lambda t_2})}{(e^{-\lambda_2 t_1} + e^{-\lambda_2 t_2})(2 - e^{-\lambda t_1} - e^{-\lambda t_2})}]. \tag{6.18}$$

From equation (6.12) we get

$$V_\Lambda(\hat{\beta}) = \frac{2}{n'[(1 - e^{-\lambda t_1})e^{-\lambda t_1} + (1 - e^{-\lambda t_2})e^{-\lambda t_2}]} +$$
$$\frac{2}{n'[(1 - e^{-\lambda_2 t_1})e^{-\lambda_2 t_1} + (1 - e^{-\lambda_2 t_2})e^{-\lambda_2 t_2}]} \tag{6.19}$$

6.2.3 The Relative Efficiency of the Logistic Model

From equations (6.16), (6.18) and (6.19) and substitution into equation (6.3), where S_1 is asymptotically $N(\beta_\Lambda, \frac{V_\Lambda(\beta)}{n})$ and S_2 is asymptotically $N(\gamma, \frac{V(\gamma)}{n})$, the relative efficiency of the two models in detecting the null hypothesis of no effect is, if $\gamma = 0$

$$e(S_1 : S_2) = \frac{4(t_1 e^{-\lambda t_1} + t_2 e^{-\lambda t_2})^2}{(2 - e^{-\lambda t_1} - e^{-\lambda t_2})^2(e^{-\lambda t_1} + e^{-\lambda t_2})^2} *$$
$$\frac{(1 - e^{-\lambda t_1})(1 - e^{-\lambda t_2})((1 - e^{-\lambda t_1})e^{-\lambda t_1} + (1 - e^{-\lambda t_2})e^{-\lambda t_2})}{t_1^2(1 - e^{-\lambda t_2})e^{-\lambda t_1} + t_2^2(1 - e^{-\lambda t_1})e^{-\lambda t_2}} \tag{6.20}$$

Equation (6.20) was evaluated for values of λt_1 and λt_2 chosen so that the probability of a patient developing active disease (in the organ/system

153

Table 6.1: The relative efficiency of the logistic model when patients are observed at two distinct time points.

Probability of failing before t_1	Probability of failing before t_2								
	10%	20%	30%	40%	50%	60%	70%	80%	90%
10%	1	0.960	0.876	0.769	0.647	0.514	0.375	0.237	0.110
20%		1	0.973	0.903	0.803	0.680	0.536	0.379	0.216
30%			1	0.977	0.913	0.814	0.683	0.522	0.339
40%				1	0.978	0.914	0.809	0.660	0.468
50%					1	0.978	0.908	0.785	0.600
60%						1	0.975	0.890	0.727
70%							1	0.968	0.847
80%								1	0.948

under consideration) before t_1 ranges from 10% to 80%, and the probability of a patient developing active disease before t_2 ranges from 10% to 90%. The results are given in table 6.1.

It can be seen from table 6.1 that when the two time points are reasonably close to each other then the relative efficiency of the logistic model to the exponential remains quite high. It is only as the probability of a patient failing between the two observation times is greater than 50% that the relative efficiency of the logistic model to the exponential falls below 50%.

6.3 Three Time Points

In order to verify that the relative efficiency of the logistic models (when not stratified by time) will decrease with the addition of extra times of observation, both groups of patients (n' in each group) were assumed to have been observed at three distinct time points. (A third of each group were observed

154

at time t_1, a third at time t_2 and a third at time t_3.) In this situation, equation (6.5) leads to

$$v(\hat{\gamma}) = \frac{3(1 - e^{-\lambda t_1})(1 - e^{-\lambda t_2})(1 - e^{-\lambda t_3})}{\lambda^2 n'[t_1^2(1 - e^{-\lambda t_2})(1 - e^{-\lambda t_3})e^{-\lambda t_1} + t_2^2(1 - e^{-\lambda t_1})(1 - e^{-\lambda t_3})e^{-\lambda t_2} + t_3^2(1 - e^{-\lambda t_1})(1 - e^{-\lambda t_2})e^{-\lambda t_3}]} *$$

$$\frac{3(1 - e^{-\lambda_2 t_1})(1 - e^{-\lambda_2 t_2})(1 - e^{-\lambda_2 t_3})}{\lambda_2^2 n'[t_1^2(1 - e^{-\lambda_2 t_2})(1 - e^{-\lambda_2 t_3})e^{-\lambda_2 t_1} + t_2^2(1 - e^{-\lambda_2 t_1})(1 - e^{-\lambda_2 t_3})e^{-\lambda_2 t_2} + t_3^2(1 - e^{-\lambda_2 t_1})(1 - e^{-\lambda_2 t_2})e^{-\lambda_2 t_3}]} \tag{6.21}$$

Equation (6.10) gives

$$\hat{\beta}_\Lambda = \log[\frac{(3 - e^{-\lambda_2 t_1} - e^{-\lambda_2 t_2} - e^{-\lambda_2 t_3})(e^{-\lambda t_1} + e^{-\lambda t_2} + e^{-\lambda t_3})}{(e^{-\lambda_2 t_1} + e^{-\lambda_2 t_2} + e^{-\lambda_2 t_3})(3 - e^{-\lambda t_1} - e^{-\lambda t_2} - e^{-\lambda_2 t_3})}]. \tag{6.22}$$

Equation (6.12) gives the relative efficiency of the two models in detecting the null hypothesis of no effect as

$$V_\Lambda(\hat{\beta}) = \frac{3}{n'[(1 - e^{-\lambda t_1})e^{-\lambda t_1} + (1 - e^{-\lambda t_2})e^{-\lambda t_2} + (1 - e^{-\lambda t_3})e^{-\lambda t_3}]} +$$

$$\frac{3}{n'[(1 - e^{-\lambda_2 t_1})e^{-\lambda_2 t_1} + (1 - e^{-\lambda_2 t_2})e^{-\lambda_2 t_2} + (1 - e^{-\lambda_2 t_3})e^{-\lambda_2 t_3}]}. \tag{6.23}$$

Substitution of equations (6.22), (6.23) and (6.21) into equation (6.3), with, as previously, S_1 asymptotically $N(\beta_\Lambda, \frac{V_\Lambda(\beta)}{n})$ and S_2 asymptotically $N(\gamma, \frac{V(\gamma)}{n})$ and, with $\gamma = 0$, gives

$$e(S_1 : S_2) = \frac{9(t_1 e^{-\lambda t_1} + t_2 e^{-\lambda t_2} + t_3 e^{-\lambda t_3})^2}{(3 - e^{-\lambda t_1} - e^{-\lambda t_2} - e^{-\lambda t_3})^2(e^{-\lambda t_1} + e^{-\lambda t_2} + e^{-\lambda t_3})^2} *$$

$$\frac{(1 - e^{-\lambda t_1})(1 - e^{-\lambda t_2})(1 - e^{-\lambda t_3})((1 - e^{-\lambda t_1})e^{-\lambda t_1} + (1 - e^{-\lambda t_2})e^{-\lambda t_2} + (1 - e^{-\lambda t_3})e^{-\lambda t_3})}{t_1^2(1 - e^{-\lambda t_2})(1 - e^{-\lambda t_3})e^{-\lambda t_1} + t_2^2(1 - e^{-\lambda t_1})(1 - e^{-\lambda t_3})e^{-\lambda t_2} + t_3^2(1 - e^{-\lambda t_1})(1 - e^{-\lambda t_2})e^{-\lambda t_3}}. \tag{6.24}$$

Equation (6.24) was then evaluated for different values of λt_1, λt_2 and λt_3. The results are given in appendix E. It can be seen that the addition of an extra time point does, as expected, reduce the relative efficiency of the logistic model to the exponential.

6.4 n' Time Points

One generalization of the results from the previous sections, derives from the assumption that the n' patients in each group are observed at n' different times $t_1, t_2....t_{n'}$. The relative efficiency of S_1 (asymptotically $N(\beta_\Lambda, \frac{V_\Lambda(\beta)}{n})$) and S_2 (asymptotically $N(\gamma, \frac{V(\gamma)}{n})$) is then

$$e(S_1 : S_2) = \frac{n'^2 \sum_{q=1}^{n'}(t_i^2 e^{-\lambda t_i})}{n' - \sum_{q=1}^{n'} e^{-\lambda t_i}} *$$
$$\frac{\prod_{q=1}^{n'}(1 - e^{-\lambda t_i}) \left(\sum_{q=1}^{n'}(1 - e^{-\lambda t_i})e^{-\lambda t_i}\right)}{\sum_{q=1}^{n'} t_i^2 e^{-\lambda t_i} \left(\prod_{w=1 w \neq q}^{n'}(1 - e^{-\lambda t_w})\right)}$$

It should be noted that this is only valid for the slightly artificial situation where the n' patients from one group are seen at the same n' distinct time points as the n' patients from the other group.

6.5 Stratification of the Logistic Model

Given that as expected the efficiency of the logistic model is reduced with the addition of each additional time of observation, it is of interest to ascertain whether the stratification of the model by time increases the relative efficiency of the logistic models to the exponential. Consider the situation described in section 6.2, and stratify the logistic model by the times t_1 and t_2. This gives a logistic model of the form

$$\log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \alpha_{t_j} + \beta x_i$$

for $i = 1, 2$ and $j = 1, 2$ where

$$\pi_{ij} = \frac{e^{(\alpha_{t_j} + \beta x_i)}}{1 + e^{(\alpha_{t_j} + \beta x_i)}}$$

is the probability of failing before time t_j for those patients in group i. This gives a likelihood of the form

$$l = \sum_{j=1}^{2}\sum_{i=1}^{2}[z_{ij}(\alpha_{t_j} + \beta x_i) - \frac{n'}{2}\log[1 + e^{(\alpha_{t_j}+\beta^* x_i)}]]$$

where z_{ij} is the number observed to fail in group i at time t_j. As previously the characteristic of interest is represented in the logistic model by a binary indicator variable x_i, with corresponding parameter β; $x_1 = 0$ for those patients in group 1, and $x_2 = 1$ for those patients in group 2. From equation (6.10) we know that

$$E_\Lambda(\frac{\partial l}{\partial \alpha_{t_j}}) = 0.$$

This gives

$$E_\Lambda \sum_{i=1}^{2}(z_{ij} - \frac{n'}{2}\pi_{ij}) = 0,$$

with

$$E_\Lambda(z_{1j}) = \frac{n'}{2}(1 - e^{-\lambda t_j})$$

$$E_\Lambda(z_{2j}) = \frac{n'}{2}(1 - e^{-\lambda_2 t_j}).$$

This leads to

$$2 - e^{-\lambda t_j} + e^{-\lambda_2 t_j} = \frac{e^{\hat{\alpha}_\Lambda t_j}}{1 + e^{\hat{\alpha}_\Lambda t_j}} - \frac{e^{(\hat{\alpha}_\Lambda t_j + \hat{\beta}_\Lambda)}}{1 + e^{(\hat{\alpha}_\Lambda t_j + \hat{\beta}_\Lambda)}}. \tag{6.25}$$

Letting

$$E_\Lambda(\frac{\partial l}{\partial \beta}) = 0$$

gives

$$E_\Lambda(\sum_{j=1}^{2}x_i(z_{ij} - \frac{n'}{2}\pi_{ij}) = 0,$$

which leads to

$$2 - e^{-\lambda_2 t_1} - e^{-\lambda_2 t_2} = \frac{e^{(\hat{\alpha}_{\Lambda 1}+\hat{\beta}_\Lambda)}}{1 + e^{(\hat{\alpha}_{\Lambda 1}+\hat{\beta}_\Lambda)}} + \frac{e^{(\hat{\alpha}_{\Lambda 2}+\hat{\beta}_\Lambda)}}{1 + e^{(\hat{\alpha}_{\Lambda 2}+\hat{\beta}_\Lambda)}}. \tag{6.26}$$

Unfortunately no closed form solution of equations (6.25) and (6.26) could be found for $\hat{\beta}$. Consequently it is not possible to calculate the Pitman relative efficiency of this model to the exponential.

6.6 The Relative Behaviour of the Stratified Logistic Model: Assessment Through Computer Simulations

In order to compare the relative behaviour of the stratified logistic model, to the exponential model, and to compare it's performance to that of the non-stratified logistic, computer simulations were used. An example of the computing used can be found in appendix A.

Two groups of 50 patients were considered, with as previously, the second group consisting of those patients with a characteristic that distinguishes them from the patients in the first group. $T_1, ..., T_{100}$ were then defined as independent random variables representing the time until the occurrence of active disease in one organ/system in each of the 100 patients. It was again assumed that $T_1, ..., T_{100}$ were exponentially distributed with

$$f(t|\lambda_i) = \lambda_i e^{-\lambda_i t} \quad t \geq 0$$

for $i = 1, 2$, and where $\lambda_1 = \lambda$ is the rate parameter for those patients in group 1 and $\lambda_2 = \lambda e^{\gamma}$ is the rate parameter for those patients in group 2. Two random samples were then generated from exponential distributions to represent the times at which the patients in each group developed active disease. However as it is only of interest to examine the relative efficiency of the logistic model to the exponential when $\gamma = 0$, both random samples were generated from the same exponential distribution with parameter λ.

For each patient it was noted whether they had developed active disease before specified times. Two binary vectors were then created indicating whether the patients in each group had 'failed' by the time of observation. Three different models, the exponential, the logistic and the stratified logistic were then used to investigate whether there appeared to be a significant difference between the two groups. Firstly the true model 'the exponential' was used. It was assumed that (as previously stated) the times to failure of the first group of patients were exponentially distributed with parameter λ and the times to failure of the second group of patients were exponentially distributed with parameter λ_2. Values of $\hat{\gamma}$ and $V(\hat{\gamma})$ were calculated from the appropriate log-likelihoods l^*, with the s-plus function 'nlmin' used to maximise the likelihood l^* setting $\frac{\partial l^*}{\partial \lambda_1} = 0$ and $\frac{\partial l^*}{\partial \lambda_2} = 0$ and equation (6.5) to calculate $V(\hat{\gamma})$. The null hypothesis of of no effect ($\hat{\gamma}=0$) was then tested using the test statistic $\frac{\hat{\gamma}}{se(\hat{\gamma})}$, which was assumed to have a N(0,1) distribution if $\gamma = 0$.

The characteristic of interest that distinguishes the second group of patients from the first was represented in all logistic and stratified logistic models by a binary indicator variable with corresponding parameter β. The pre-programmed generalized linear models function in S-plus was used to fit a logistic model and a stratified logistic model in order to estimate $\hat{\beta}$ and $V(\hat{\beta})$. In each case the null hypothesis of no effect ($\hat{\beta}=0$) was tested using the test statistic $\frac{\hat{\beta}}{se(\hat{\beta})}$, which was assumed to have a N(0,1) distribution if $\beta = 0$.

A 5% level of significance was used and the simulations were run 1000 times.

The times of observation were chosen so the probability of a patient developing active disease before being observed ranged from 10% to 90%. The value of λ chosen was 0.02 as this corresponds roughly to the observed fail-

159

ure rate for developing active mucocutaneous disease. The times therefore ranged from 5 to 155 days.

6.6.1 One time point

In order to establish the usefulness of running the simulation study, initial simulations were run where all patients were observed at one time point t. The results of theses simulations are given with the type 1 error rates for each of the models.

Table 6.2: Type 1 error rates for the exponential and logistic models when the simulations were run with all patients being observed at the same time

$1 - e^{-\lambda t}$	Type 1 error rate	
	Exponential model	Logistic model
10%	0.023	0.034
20%	0.049	0.049
30%	0.044	0.044
40%	0.062	0.062
50%	0.058	0.058
60%	0.054	0.054
70%	0.057	0.054
80%	0.059	0.047

Table 6.2 shows that for one time point the behaviour of the two approaches is comparable, with some finite sample variation from the 5% type 1 error rate at either end. For all the simulations where $20\% \leq 1 - e^{-\lambda t} \leq 60\%$, the type 1 error rates are equal.

6.6.2 Two time points

It was then assumed that half the patients in each group were observed at time t_1 the others being observed at time t_2. Simulations were run for values of t_1 and t_2 such that $20\% \leq 1 - e^{-\lambda t_1} \leq 60\%$ and $30\% \leq 1 - e^{-\lambda t_2} \leq 70\%$.

Table 6.3: Type 1 error rates for the exponential, logistic and stratified logistic models when the simulations were run with half the patients in each group being observed at t_1 and the other half at time t_2

$1 - e^{-\lambda t_1}$	$1 - e^{-\lambda t_2}$	Type 1 error rate		
		Exponential model	Logistic model	Stratified Logistic Model
20%	70%	0.063	0.033	0.062
20%	60%	0.048	0.031	0.045
20%	50%	0.064	0.051	0.064
20%	40%	0.055	0.044	0.060
20%	30%	0.040	0.038	0.041
30%	70%	0.056	0.045	0.055
30%	60%	0.053	0.047	0.058
30%	50%	0.045	0.045	0.046
30%	40%	0.051	0.048	0.050
40%	70%	0.042	0.034	0.041
40%	60%	0.050	0.048	0.050
40%	50%	0.062	0.067	0.067
50%	70%	0.054	0.052	0.053
50%	60%	0.069	0.070	0.070
60%	70%	0.053	0.049	0.050

It is clear from table 6.3 that generally the type 1 error rate of the stratified logistic model is closer to the type 1 error rate of the exponential model than is the type 1 error rate of the simple logistic model. This shows com-

161

parable behaviour for the stratified logistic model and the exponential, and demonstrates that there is a problem if the time factor is not adequately modelled (the simple logistic model). The difference between the results gets less as the time points get closer, as would be expected.

6.6.3 Multiple Time Points

In an attempt to assess the relative behaviour of simple logistic models and the models used in chapter 5 (logit $\Pr(y_{it} = 1 \mid y_{it-1} = 0) = \alpha_0 + \beta x_{it} + \delta(t'_{ip} - (t'_{ip-1})))$ to the exponential, simulations were run where the times were allowed to vary normally around two time points with a standard deviation corresponding to $\lambda t = 0.2$ (this corresponds roughly to a standard deviation of about 10 days). The first mean time point was fixed so that $1 - e^{-\lambda t_1} = 40\%$ and the simulations were run for mean values of t_2 so that $50\% \leq 1 - e^{-\lambda t_2} \leq 70\%$.

Table 6.4: Type 1 error rates for the exponential, logistic and logistic models adjusted for time between visits when the simulations were run with the times of observation for the patients in each group varying normally around two time points

		Type 1 error rate		
$1 - e^{-\lambda t_1}$	$1 - e^{-\lambda t_2}$	Exponential model	Logistic model	Stratified Logistic Model
40%	70%	0.060	0.045	0.059
40%	60%	0.050	0.058	0.051
40%	50%	0.054	0.055	0.056

Table 6.4 indicates that when there are a large number of time points, the type 1 error rate of the logistic model adjusted for time between visits is

162

generally closer to the type 1 error rate of the exponential model than is the type 1 error rate of simple logistic model (as was the case for two distinct time points). These results demonstrate that 'empirical modelling of the time effect' is reasonably effective in getting the size of the test comparable to the exponential

6.6.4 Conclusions

The models used in chapter 5 are particularly appropriate if the time between the visits does not affect a patient's chance of presenting with active disease. Dynamic covariates were used without any appeal to an underlying model in continuous time. However the results of this chapter indicate that not only is little efficiency lost relative to an underlying model if the times between the visits are relatively similar, but in addition that if there is an underlying model the logistic models used are still useful if they are either stratified by time or adjustment is made for the time since the previous visit.

Chapter 7

Discussion

Chapters 3 and 4 describe new statistical methodologies that were developed for studies in two areas of current research in rheumatology: myositis and lupus. The development of measures for the assessment of disease activity and damage in myositis, based on previous work in lupus, led to the identification of a need for an approach to the assessment of the reliability of the tools. The existence of a unique 10 year data set of measurements of disease activity and damage in lupus patients led to the development of an approach based on the analysis of multinomial panel data.

The statistical work presented in this thesis extends available statistical techniques and their range of application. In addition, without the work on agreement and reliability it would not have been possible to comment on the performance of the newly developed myositis measures and the progress towards the current validation exercises would have been compromised. It is also hoped that the results of the analysis of the lupus data will enable physicians to better understand the pattern of disease presentation in lupus.

7.1 Myositis

A number of issues became apparent as the work on the development of the myositis tools proceeded. In particular the large number of measures, with different measurement scales, highlighted the need for simple statistics that could summarize the performance of all the tools.

As an initial approach to the analysis of the performance of the tools in the first real patient exercise, the analysis of variance for each tool was examined and the percentages of variance attributable to each source of error were calculated and presented to the participating physicians. However, for the physicians, the large number of outcomes and the different measurement scales led to some confusion with the interpretation of these results.

Further examination of available literature (Shoukri 2004) highlighted a lack of clarity on the issues of reliability and agreement, and no recognition of the need to address both issues when evaluating the performance of tools, such as those developed for myositis. It became clear that no single measure existed that would adequately sum up both the reliability of a measure, and the level of physician agreement in a test-retest exercise.

As an initial step in addressing the need for a suitable measure, an intraclass correlation coefficient, suitable for the Latin square experimental design, was devised. As discussed, despite the distributional properties being best understood for fully continuous measures, an intraclass correlation coefficient can reasonably be used for all the outcomes being examined. It became apparent, through examination of the results, that the intraclass correlation coefficient is a measure of reliability despite frequently being described (Shrout and Fleiss 1979) as a measure of agreement. The myositis patients participating in the exercises were chosen for their clinical diversity, and yet given the number of areas being assessed, it was inevitable that there

was little activity and/or damage in some organs/systems. This lack of variation led to low intraclass correlation coefficients for some organs/systems. It is well established that it is difficult to achieve high reliability of any measure in very homogeneous populations because it is difficult to distinguish between patients in a population where the differences are either rare or very fine. The low intraclass correlation coefficients for the organs/systems where the patients were very homogeneous emphasised that, in order to adequately assess the performance of the tools, it would be necessary to also measure the level of physician agreement.

A number of statistics have been suggested as measures of agreement. Generally kappa and weighted kappa are described as measures of agreement for nominal and ordinal data respectively (Kraemer *et al.* 2002) and a concordance correlation coefficient (Lin 1989) is described as a measure of agreement for continuous outcomes. However weighted kappa has been shown to be equivalent to the intraclass correlation coefficient when quadratic weights are used (Fleiss and Cohen 1973) and the concordance correlation coefficient has been shown to be equivalent to an intraclass correlation coefficient from a two way analysis of variance (Nickerson 1997).

The lack of a suitable existing measure of agreement led to the development of the new measure r. Then, in order to facilitate the interpretation of the combination of the measures of reliability and agreement, and to provide a single measure that adequately describes the performance of the tools, the classification suggested in chapter 3 was devised.

Chapter 3 thus presents an new approach to the assessment of reliability and agreement. The results presented provide a clear summary of the performance of a large number of newly developed tools. The participating physicians found the results easy to interpret and these new measures have

since been used in assessing the performance of a new version of the BILAG index.

7.2 Lupus

The existence of a unique data set, with data from a large number of patients over a time period of 10 years, led to the opportunity to undertake the analysis described in chapters 4 and 5. The work done represents the first development of models for disease activity in lupus that can reflect the multifaceted nature of the disease.

The motivation for the analyses was an attempt to validate the existing clinical impression that subsets of lupus exist within the disease, and an attempt to provide physicians with reliable indicators of future disease activity. The focus of the analysis was the investigation of interrelationships between disease activity in the different organs/systems.

The complex nature of the data and the lack of the exact timing of events, led to difficulties with the analysis of the data in continuous time.

The approach taken was to model a patient's state at a clinic visit. The use of simple logistic regression models to estimate the parameters of the polychotomous logistic model enabled the fitting of complex models that would otherwise have been hindered by the lack of flexibility of the current computer packages. The approach taken also allowed for the models to be fitted using generalized estimating equations with an exchangeable correlation structure, to adjust for any dependence not captured by conditioning on a patient's previous state. The use of dynamic covariates permitted the modelling of a patient's disease history without recourse to models in continuous time which are particularly problematic with panel data and time

dependent explanatory variables.

An area of particular concern, in the development of the approach, was the irregular time intervals between visits. It is possible that a patient's chance of presenting at a visit could be highly dependent on the time since the previous visit. If true this would throw doubt on the suitability of the approach taken and require further investigation into the effect of the time between visits. However, it was shown that this was not the case and, therefore, the use of separate logistic regressions with dynamic covariates for the analysis of these multinomial panel data is supported.

In addition to this empirical investigation of the appropriateness of the modelling approach, more general consideration of the efficiency of the approach relative to the use of models in continuous time may provide an indication of its wider applicability. Chapter 6 describes an investigation into the relative efficiency of the logistic models when the underlying distribution of the time to disease activity is exponential. Pitman efficiency was used to measure the relative efficiency of logistic models to an exponential model in testing the null hypothesis of no regression effect. This is a simple situation and the results are asymptotic, however the results should give qualitative guidance of the relative behaviour of the two models.

The method confirmed, as expected, that if all patients were seen at the same time intervals a logistic model is as efficient as the exponential in testing the null hypothesis. Naive logistic models (not stratified by time) were then compared to the exponential model when the patients were seen at different time intervals. As the number of time intervals increased the relative efficiency of the logistic models decreased. It was interesting to note however, that relatively little efficiency was lost if the time intervals did not differ too greatly.

Computer simulations were used to attempt to assess the relative efficiency of logistic models, stratified by time, to the exponential model. The results of the simulations indicated that if there is an underlying continuous distribution the logistic models used are still appropriate if stratified by time or if adjustment is made for the time between visits.

The conclusions on the relative efficiency of the logistic models are based on the assumption that the true model is known. They thus provide a conservative evaluation of the performance of the logistic models in general.

From a medical perspective the results do suggest that subsets of lupus exist. It was hypothesized that this knowledge might benefit physicians when deciding on appropriate treatment but this remains an open question.

Appendix A

Computer programs

A.1 Analysis of the Myositis Exercises

```
blooms2_scan("a:blooms22.txt",list(Bphysician=0,Btime=0,Bpatient=0))
attach(blooms2)
exp<-cbind(Bphysician,Bpatient)
mat<-matrix(0,6,7)
mspatients<-c(rep(0,43))
msphysicians<-c(rep(0,43))
mstime<-c(rep(0,43))
msresiduals<-c(rep(0,43))
options(contrasts=c("contr.sum","contr.sum"))
act<-scan("a:dat2.txt",list(MITAXConstitutional=0,MITAXMucocutaneous=0,
MITAXSkeletal=0,MITAXGastrointestinal=0,MITAXPulmonary=0,
MITAXCardiac=0,MITAXMuscle=0,MyoactConstitutional=0,
MyoactMucocutaneous=0,MyoactSkeletal=0,MyoactGastrointestinal=0,
```

```
MyoactPulmonary=0,MyoactCardiac=0,MyoactMuscle=0,

PhysiciansGlobal=0,Physiciansextramuscular=0,MDIMuscle=0,MDISkeletal=0,

MDICutaneous=0,MDIGastrointestinal=0,MDIpulmonary=0,MDICV=0,MDIPV=0,

MDIendocrine=0,MDIocular=0,MDIinfection=0,MDImalignancy=0,MDIother=0,

MDIglobal=0,myodamMuscle=0,

myodamSkeletal=0,myodamCutaneous=0,myodamGastrointestinal=0,

myodampulmonary=0,myodamCV=0,myodamPV=0,myodamendocrine=0,

myodamocular=0,myodaminfection=0,myodammalignancy=0,myodamother=0,

myodamglobal=0))

attach(act)

code_c(0,0,1,3,9)

Bcodeact1_code[MITAXConstitutional]

Bcodeact2_code[MITAXMucocutaneous]

Bcodeact3_code[MITAXSkeletal]

Bcodeact4_code[MITAXGastrointestinal]

Bcodeact5_code[MITAXPulmonary]

Bcodeact6_code[MITAXCardiac]

Bcodeact7_code[MITAXMuscle]

N<-6

I<-7

errordf<-24

patientdf<-5

physdf<-6

s<-I/N

rat<-errordf/patientdf
```

```
Bcodeacttotal_Bcodeact1+Bcodeact2+Bcodeact3+Bcodeact4+Bcodeact5+Bcodeact6+
Bcodeact7
Bactivity_as.matrix(rbind(Bcodeact1,Bcodeact2,Bcodeact3,Bcodeact4,Bcodeact
Bcodeact6,Bcodeact7,Bcodeacttotal,MyoactConstitutional,MyoactMucocutaneous
MyoactSkeletal,MyoactGastrointestinal,MyoactPulmonary,MyoactCardiac,
MyoactMuscle,PhysiciansGlobal,Physiciansextramuscular,MDIMuscle,MDISkeleta
MDICutaneous,MDIGastrointestinal,MDIpulmonary,MDICV,MDIPV,MDIendocrine,
MDIocular,MDIinfection,MDImalignancy,MDIother,
MDIglobal,myodamMuscle,myodamSkeletal,myodamCutaneous,
myodamGastrointestinal,myodampulmonary,myodamCV,myodamPV,myodamendocrine,
myodamocular,myodaminfection,
myodammalignancy,myodamother,myodamglobal))
attach(act)
code_c(0,0,1,3,9)
FBphysician_as.factor(Bphysician)
FBpatient_as.factor(Bpatient)
FBtime_as.factor(Btime)
patient<-c(rep(0,43))
physician<-c(rep(0,43))
time<-c(rep(0,43))
measure<-c("MITAX Constitutional","MITAX Mucocutaneous","MITAX Skeletal",
"MITAX Gastrointestinal","MITAX Pulmonary","MITAX Cardiac","MITAX Muscle",
"total MITAX","Myoact Constitutional","Myoact Mucocutaneous",
"Myoact Skeletal","Myoact Gastrointestinal","Myoact Pulmonary",
"Myoact Cardiac","Myoact Muscle","Physicians Global",
"Physicians extra muscular","MDI Muscle","MDI Skeletal","MDI Cutaneous",
"MDI Gastrointestinal","MDI pulmonary","MDI CV","MDI PV","MDI endocrine",
```

"MDI ocular","MDI infection","MDI malignancy","MDI other","MDI global",

"Myodam Muscle","Myodam Skeletal","Myodam Cutaneous",

"Myodam Gastrointestinal","Myodam Pulmonary","Myodam CV","Myodam PV",

"Myodam Endocrine","Myodam Ocular","Myodam Infection","Myodam Malignancy"

"Myodam Other","Myodam Global")

```
for (i in 1:43){lm1_lm(Bactivity[i,]~FBpatient+FBphysician+FBtime)
a<-anova(lm1)
b<-a$"Mean Sq"
patient[i]<-round((b[1]/(b[1]+b[2]+b[3]+b[4]))*100,2)
physician[i]<-round((b[2]/(b[1]+b[2]+b[3]+b[4]))*100,2)
time[i]<-round((b[3]/(b[1]+b[2]+b[3]+b[4]))*100,2)
mspatients<-round(b[1],2)
msphysicians<-round(b[2],2)
mstime<-round(b[3],2)
msresiduals<-round(b[4],2)
print(anova(lm1))
for (j in 1:42){d_as.vector(exp[j,1])
e<-as.vector(exp[j,2])
mat[e,d]_Bactivity[i,j]}
print(measure[i])
print(mat)

par(mfrow=c(2,1))
hist(resid(lm1),xlab="residual")
qqnorm(resid(lm1),xlab="theoretical",ylab="empirical")
sigmapatient<-sqrt((mspatients-msresiduals)/N)
```

```
sigmaphysician<-sqrt((msphysicians-msresiduals)/I)
r<-sigmaphysician/sigmapatient


rho<-N*(mspatients-msresiduals)/
(N*mspatients+I*msphysicians+(N*I-I-N)*msresiduals)


Fphys<-msphysicians/msresiduals


top<-errordf*(I*rho*Fphys+N+(N*I-N-I)*rho)^2
bottom<-(errordf/physdf)*I^2*rho^2*Fphys^2+(N+(N*I-N-I)*rho)^2
nu<-top/bottom
Fstar<-qf(0.975,patientdf,nu)
Fstar2<-qf(0.975,nu,patientdf)
lowertop<-N*(mspatients-Fstar*msresiduals)
lowerbottom<-Fstar*(I*msphysicians+(N*I-N-I)*msresiduals)+N*mspatients
lower<-lowertop/lowerbottom
uppertop<-N*(Fstar2*mspatients-msresiduals)
upperbottom<-Fstar2*N*mspatients+(N*I-N-I)*msresiduals+I*msphysicians
upper<-uppertop/upperbottom


top2<-errordf*(r^2*mspatients+(1-s*r^2)*msresiduals)^2
bottom2<-rat*s*r^4*mspatients^2+(1-s*r^2)^2*msresiduals^2
nu2<-top2/bottom2


newf<-qf(0.975,physdf,nu2)
```

174

```
newf2<-qf(0.975,nu2,physdf)


rlower<-sqrt(N*(msphysicians-newf*msresiduals)/

(I*newf*(mspatients-msresiduals)))

rupper<-sqrt(N*(newf2*msphysicians-msresiduals)/

(I*(mspatients-msresiduals)))

print("ICC")

print(round(rho,3))

CI<-cbind(round(lower,3),round(upper,3))

print(CI)

print("se patient")

print(round(sigmapatient,3))

print("se physician")

print(round(sigmaphysician,3))

print("se patient/se physician")

print(round(r,3))


CI2<-cbind(round(rlower,3),round(rupper,3))


print(CI2)}
```

A.2 Efficiency simulation: multiple time points

```
sig2<-c(rep(0,1000))

sig3<-c(rep(0,1000))
```

```
sig4<-c(rep(0,1000))

library(gee)

id<-c(1:100)

coef<-c(rep(0,1000))

coefexp<-c(rep(0,1000))

betastrat<-c(rep(0,1000))

varx<-c(rep(0,1000))

varxexp<-c(rep(0,1000))

varstrat<-c(rep(0,1000))

lambda<-0.02

tt1_-log(0.4)/lambda

tt2_-log(0.2)/lambda


for (j in 1:1000){

vec1<-rexp(50,lambda)

vec2<-rexp(50,lambda)

logistic1<-c(rep(0,50))

logistic2<-c(rep(0,50))

x<-c(rep(0,50),rep(1,50))

t1<-round(c(rnorm(25,tt1,10),rnorm(25,tt2,10)),0)

t2<-round(c(rnorm(25,tt1,10),rnorm(25,tt2,10)),0)

t<-as.vector(cbind(t1,t2))

for (i in 1:50){

if(vec1[i]<=t1[i]) {logistic1[i]<-1}  else {logistic1[i]<-0}

if(vec2[i]<=t2[i]) {logistic2[i]<-1}  else {logistic2[i]<-0}}

t11<-numeric(0)
```

```
for(i in 1:50)
if(logistic1[i]==1)
t11<-c(t11,t1[i])
t11

t12<-numeric(0)
for(i in 1:50)
if(logistic1[i]==0)
t12<-c(t12,t1[i])
t12

t21<-numeric(0)
for(i in 1:50)
if(logistic2[i]==1)
t21<-c(t21,t2[i])
t21

t22<-numeric(0)
for(i in 1:50)
if(logistic2[i]==0)
t22<-c(t22,t2[i])
t22

a<-length(t11)
b<-length(t12)
c<-length(t21)
d<-length(t22)
```

```
func1<-function(x)
{

x1 <-  - (t11 * x)
x2 <- 1 - exp(x1)
v <-   log(x2)
v2<-t12*x
-sum(v) +sum(v2)
}
 func2<-function(x)
{

x1 <-  - (t21 * x)
x2 <- 1 - exp(x1)
v <-   log(x2)
v2<-t22*x
-sum(v) +sum(v2)
}

logistic<-as.vector(cbind(logistic1,logistic2))

lambda1<-nlmin(func1,1)
lambda2<-nlmin(func2,1)

beta<-log(lambda2$x/lambda1$x)
```

```
model<-glm(logistic~x,family=binomial)
modelx<-glm(logistic~x+t,family=binomial)


model1<-gee(logistic~x,id,family=binomial)
model2<-gee(logistic~x+t,id,family=binomial)


varstrat[j]<-model2$robust.variance[2,2]
betastrat[j]<-modelx$coef[2]
sig5<-betastrat[j]/(sqrt(varstrat[j]))
sig55<-abs(sig5)
if(sig55>1.96)
sig4[j]<-1 else sig4[j]<-0


coef[j]<-model$coef[2]
coefexp[j]<-beta
varx[j]<-model1$robust.variance[2,2]
sig<-coef[j]/sqrt(varx[j])
modsig<-abs(sig)
if(modsig >1.96)
sig2[j]<-1 else sig2[j]<-0


x<-lambda1$x
y<-lambda2$x


var1<-c(rep(0,a))
for (i in 1:a){
```

```
x1 <-  - (t11[i] * x)
x2 <- 1 - exp(x1)
x3<-x2*x2
x4<-t11[i]*t11[i]
x5<-x4*exp(x1)
x6<-x5/x3
var1[i]<-x6}


part1<-sum(var1)


var2<-c(rep(0,c))
for (i in 1:c){
x1 <-  - (t21[i] * y)
x2 <- 1 - exp(x1)
x3<-x2*x2
x4<-t21[i]*t21[i]
x5<-x4*exp(x1)
x6<-x5/x3
var2[i]<-x6}


part2<-sum(var2)


part3<-1/(x^2*part1)
part4<-1/(y^2*part2)


varxexp[j]<-part3+part4
```

```
sig1<-beta/(sqrt(varxexp[j]))
sig11<-abs(sig1)
if(sig11>1.96)
sig3[j]<-1 else sig3[j]<-0

print(j)}
betahat<-sum(coef)/1000
betahatexp<-sum(coefexp)/1000
betahatstrat<-sum(betastrat)/1000
siglog<-sum(sig2)
sigexp<-sum(sig3)
sigstrat<-sum(sig4)
varmle<-var(coef)
varmlexp<-var(coefexp)
varstratxx<-var(betastrat)
varxx<-sum(varx)/1000
varxxexp<-sum(varxexp)/1000
varxstrat<-sum(varstrat)/1000



emselog<-sum(coef*coef)/1000
emseexp<-sum(coefexp*coefexp)/1000
emsestrat<-sum(betastrat*betastrat)/1000
testlog<-coef/(sqrt(varx))
meantestlog<-mean(testlog)
```

```
vartestlog<-var(testlog)

testexp<-coefexp/(sqrt(varxexp))

meantestexp<-mean(testexp)

vartestexp<-var(testexp)

teststrat<-betastrat/(sqrt(varstrat))

meanteststrat<-mean(teststrat)

varteststrat<-var(teststrat)


d<-c(rep(0,1000))

e<-c(rep(0,1000))

f<-c(rep(0,1000))

g<-c(rep(0,1000))

for(j in 1:1000){

if(sig2[j]==0&&sig3[j]==0){d[j]<-1}else{d[j]<-0}

if(sig2[j]==0&&sig3[j]==1){e[j]<-1}else{e[j]<-0}

if(sig2[j]==1&&sig3[j]==0){f[j]<-1}else{f[j]<-0}

if(sig2[j]==1&&sig3[j]==1){g[j]<-1}else{g[j]<-0}

}


dd<-sum(d)

ee<-sum(e)

ff<-sum(f)

gg<-sum(g)

k<-cbind(dd,ee)

l<-cbind(ff,gg)

tab<-rbind(k,l)
```

```
h<-c(rep(0,1000))
k<-c(rep(0,1000))
m<-c(rep(0,1000))
n<-c(rep(0,1000))


for(j in 1:1000){
if(sig4[j]==0&&sig3[j]==0){h[j]<-1}else{h[j]<-0}
if(sig4[j]==0&&sig3[j]==1){k[j]<-1}else{k[j]<-0}
if(sig4[j]==1&&sig3[j]==0){m[j]<-1}else{m[j]<-0}
if(sig4[j]==1&&sig3[j]==1){n[j]<-1}else{n[j]<-0}
}


hh<-sum(h)
kk<-sum(k)
mm<-sum(m)
nn<-sum(n)
o<-cbind(hh,kk)
p<-cbind(mm,nn)
tab2<-rbind(o,p)
print(tt1)
print(tt2)
print(tab)


print(tab2)


print(emseexp)
```

```
print(emselog)
print(emsestrat)
print(meantestexp)
print(meantestlog)
print(meanteststrat)
print(vartestexp)
print(vartestlog)
print(varteststrat)
```

Appendix B

Examples of graphs used to assess the normality of the residuals in the myositis experiments

Figure B.1: Histogram and qq plot of residuals from the analysis of variance of the observations of the constitutional element of the MITAX from the second patient exercise
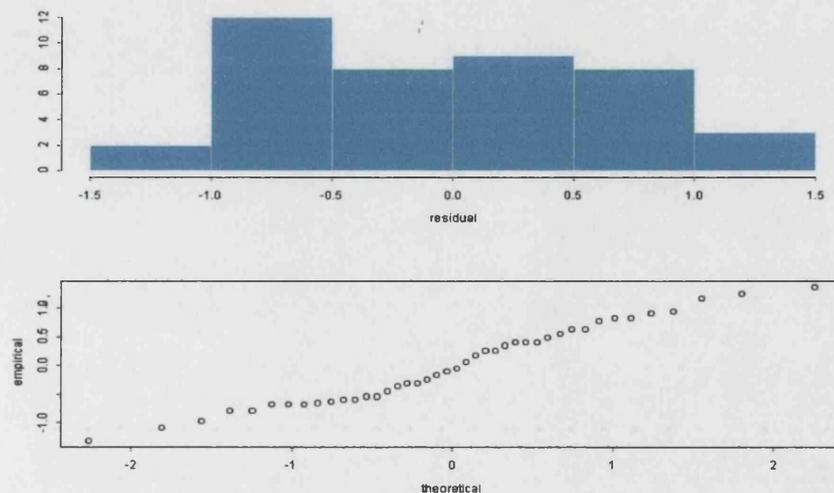
Figure B.2: Histogram and qq plot of residuals from the analysis of variance of the observations of the constitutional element of the MYOACT from the second patient exercise
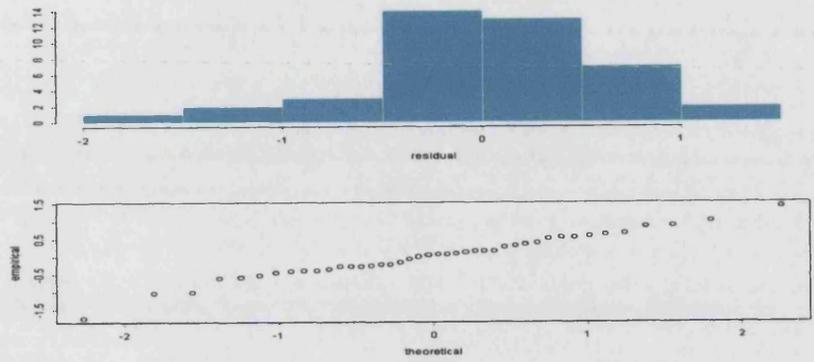
Figure B.3: Histogram and qq plot of residuals from the analysis of variance of the observations of the cutaneous element of the MDI from the second patient exercise
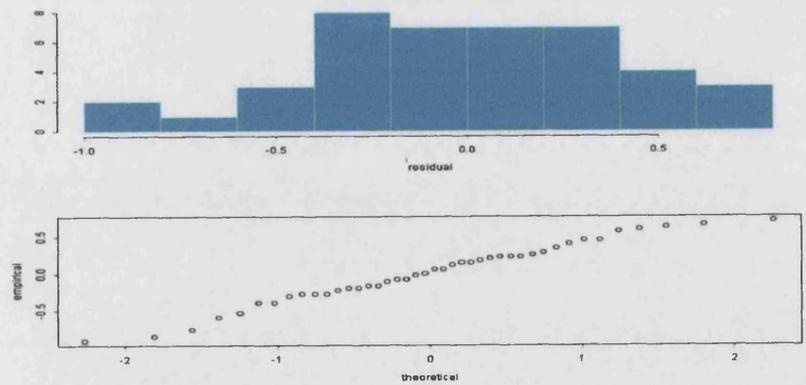
Figure B.4: Histogram and qq plot of residuals from the analysis of variance of the observations of the skeletal element of the MYODAM from the second patient exercise
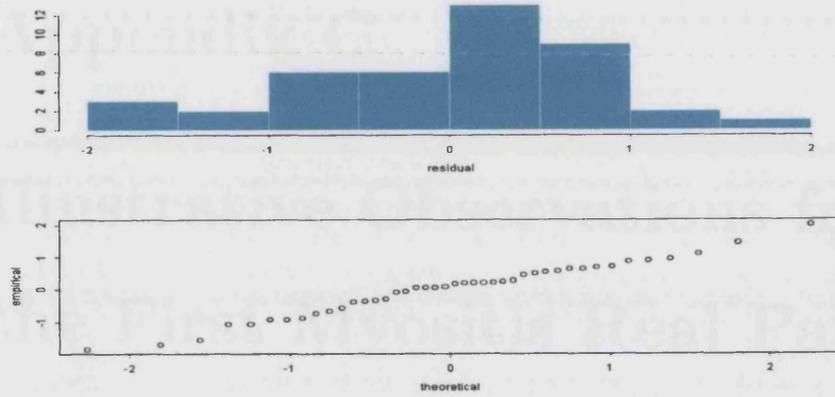
Appendix C

Illustrative Observations from the First Myositis Real Patient Exercise

The observations for a tool that differentiates well between patients and where the level of physician agreement was high.

Manual Muscle QUADRICEPS

patient	physician						
	1	2	3	4	5	6	7
1	6	7	5	2	6	3	5
2	9	7	9	8	9	8	8
3	9	10	10	10	10	10	10
4	10	9	10	10	10	10	10
5	10	10	10	10	9	8	10
6	7	7	6	6	6	6	6
7	10	10	10	10	10	10	10

The observations for a tool that appears not to differentiate well between patients and where the level of physician agreement was poor.

MITAX Muscle Disease

patient	physician						
	1	2	3	4	5	6	7
1	3	3	3	9	3	0	3
2	3	9	9	9	9	9	9
3	3	1	0	9	3	0	3
4	3	1	3	9	3	0	3
5	9	9	9	9	9	3	3
6	3	3	3	3	3	0	3
7	0	1	3	3	3	1	3

The observations for a tool that appears to differentiate well between patients but where the level of physician agreement is low.

Manual Muscle WRIST

patient	physician						
	1	2	3	4	5	6	7
1	9	9	8	6	9	9	10
2	10	9	10	7	9	9	10
3	10	10	10	8	10	10	10
4	10	10	10	7	10	9	10
5	7	6	6	5	7	5	7
6	9	10	9	8	10	8	10
7	10	9	10	9	9	10	10

The observations for a tool where a homogeneous population means a poor ability in differentiating between patients but the level of physician agreement is high.

MITAX Skeletal Disease

patient	physician						
	1	2	3	4	5	6	7
1	0	0	0	0	0	0	0
2	3	1	1	3	1	3	1
3	0	0	0	0	0	0	0
4	0	1	1	3	3	0	0
5	3	3	0	9	3	9	3
6	0	0	0	0	0	0	0
7	0	1	3	3	3	3	1

Appendix D

Details of the Analysis of the Lupus Data

D.1 The Mucocutaneous Organ/System

D.1.1 Univariate Analyses

Initial univariate analyses were carried out, the results of which are given with unadjusted odds ratios in table 5.3.

Table D.1: Unadjusted odds ratios for each additional mucocutaneous B score.

Number of mucocutaneous B's scored by the previous visit	Odds ratios for a mucocutaneous B score
0	1.647
1	1.496
2	1.359
3	1.235
4	1.122
5	1.019
6	0.926
7	0.841

From table 5.3 it can be seen that a history of either mucocutaneous A scores or mucocutaneous B scores increases a patient's chance of presenting with active mucocutaneous disease. For both A and B scores both variables are significant.

The effects of both the number of mucocutaneous A scores and the number of mucocutaneous B scores appear not to be linear and consequently quadratic factors have been included for both. This shows that although the chance of presenting with active mucocutaneous disease increases with each additional observation of either a mucocutaneous A or B, the effect of each additional observation decreases for each subsequent observation. To demon-

strate this the odds ratios for each additional observation of a mucocutaneous B are given in table D.1.

A history of renal A scores significantly affects a patient's chance of presenting with active mucocutaneous disease, with both variables giving significant results. The coefficient of a previous occurrence of a renal A indicates that having a history of renal A scores decreases a patient's chance of presenting with active mucocutaneous disease. However the number of renal A scores appears not to be linear, and a quadratic factor has been included, showing that the effect changes with each additional observation of a renal A, as shown in table D.2.

Table D.2: Unadjusted odds ratios for each additional renal A score.

Number of renal A's scored by the previous visit	Odds ratios for a renal A score
0	0.296
1	1.133
2	4.345

A history of renal B scores does not significantly affect a patient's chance of presenting with active mucocutaneous disease, with neither variable being significant.

Neither a history of musculoskeletal A scores nor a history of musculoskeletal B scores significantly affects a patient's chance of presenting with active mucocutaneous disease. In both cases neither variable is significant.

Time since the fist clinic visit, time since the previous visit and the time since the patient was last observed with active mucocutaneous disease all have no significant effect on a patient's chance of presenting with active mucocutaneous disease.

193

History of disease activity in the organs/systems not chosen for detailed analysis does not significantly affect a patient's chance of presenting with active mucocutaneous disease.

Mucocutaneous damage significantly increases a patient's chance of presenting with active disease. Neither renal nor musculoskeletal damage has a significant effect.

D.1.2   Analysis by individual organs/systems

The six variables representing history of disease activity in the mucocutaneous organ/system, together lead to a change in deviance of 96.774 that is significant on χ_6^2 ($p \approx 0$).

The number of mucocutaneous a scores and the number of mucocutaneous B scores have both lost significance on being included in the model with a previous occurrence of a mucocutaneous A and a previous occurrence of a mucocutaneous B respectively. The 4 variables lead to a change in deviance of 3.68 that is not significant on χ_4^2 ($p = 0.451$). It appears therefore that in this analysis a history of mucocutaneous activity is adequately represented by a previous occurrence of a mucocutaneous B and a previous occurrence of a mucocutaneous A as shown on table D.3. Both variables together lead to a change in deviance of 93.094 that is significant on χ_2^2 ($p \approx 0$).

The five variables representing a history of renal activity as shown in table D.4, together lead to a change in deviance of 12.736 that is significant on χ_5^2 ($p = 0.026$). However in this analysis as in the univariate analyses a history of renal B scores does not significantly affect a patient's chance of presenting with active mucocutaneous disease, both variables together lead to a change in deviance of 0.199 that is not significant on χ_2^2 ($p = 0.905$). A previous occurrence of a renal A has also lost significance it leads to a change

Table D.3: Coefficients with adjusted odds ratios from the analysis by individual organ/system (mucocutaneous) of active vs. not active mucocutaneous disease.

	coefficient	odds ratio
A previous occurrence of a mucocutaneous B	0.956	2.602
A previous occurrence of a mucocutaneous A	0.965	2.626

Table D.4: Coefficients with adjusted odds ratios from the analysis by individual organ/system (renal) of active mucocutaneous disease vs. inactive mucocutaneous, musculoskeletal and renal disease.

	coefficient	odds ratio
A previous occurrence of a renal B	0.115	1.122
Number of renal B scores	-0.019	0.981
A previous occurrence of a renal A	11.640	NA
Number of renal A scores	-17.160	
(Number of renal A scores)^2	4.478	NA

in deviance of 2.101 that is not significant on χ_1^2 ($p = 0.147$), and leaves in this analysis, the number of renal A's adequately representing all history of renal activity.

The four variables shown in table D.5 representing a history of musculoskeletal disease activity together lead to a change in deviance of 7.350 that is not significant on χ_4^2 ($p = 0.119$). None of the variables significantly affect a patient's chance of presenting with active mucocutaneous disease.

D.1.3 Multivariate Analysis

All variables were now included in a single model with only main effects.

Table D.5: Coefficients with adjusted odds ratios from the analysis by individual organ/system (musculoskeletal) of active mucocutaneous disease vs. inactive mucocutaneous, musculoskeletal and renal disease.

	coefficient	odds ratio
A previous occurrence of a musculoskeletal B	-0.219	0.804
Number of musculoskeletal B scores	0.117	1.125
A previous occurrence of a musculoskeletal A	-0.505	0.603
Number of musculoskeletal A scores	0.005	1.005

A history of disease activity in the five organs/systems not chosen for detailed analysis continues not to significantly affect a patient's chance of presenting with active mucocutaneous disease. It leads to a change in deviance of 0.171 that is not significant on χ_1^2 ($p = 0.679$).

A history of musculoskeletal B scores continues to have no significant effect on a patient's chance of presenting with active mucocutaneous disease. It leads to a change in deviance of 0.807 that is not significant on χ_2^2 ($p = 0.668$).

A history of musculoskeletal A scores has become significant on being included in the model with a history of mucocutaneous B scores. Both variables together lead to a change in deviance of 7.779 that is significant on χ_2^2 ($p = 0.020$). Table D.6 shows that the effect of a history of musculoskeletal A scores is clearer once a history of mucocutaneous B scores has been adjusted for. However a previous occurrence of a musculoskeletal A leads to a change in deviance of 0.079 that is not significant on χ_1^2 ($p = 0.779$), and the number of musculoskeletal A scores leads to a change in deviance of 2.775 that is not significant on χ_1^2 ($p = 0.096$). This indicates that in this analysis a history of musculoskeletal A scores is adequately represented by either variable. For

simplicity a previous occurrence of a musculoskeletal A score will be used.

Table D.6: The % of visits with details of patients' histories of musculoskeletal A scores and mucocutaneous B scores.

No history of musculoskeletal A scores	History of musculoskeletal A scores
$\frac{212}{2710} = 7.82\%$	$\frac{21}{354} = 5.93\%$

	No history of musculoskeletal A scores	History of musculoskeletal A scores
No history of mucocutaneous B scores	$\frac{81}{1690} = 4.79\%$	$\frac{2}{162} = 1.23\%$
History of mucocutaneous B scores	$\frac{131}{1020} = 12.84\%$	$\frac{19}{192} = 9.90\%$

A history of renal B scores leads to a change in deviance of 1.994 that is not significant on χ_2^2 ($p = 0.369$).

A history of renal A scores leads to a change in deviance of 15.878 that is significant on χ_3^2 ($p = 0.001$). However the number of renal A scores leads to a change in deviance of 4.110 that is not significant on χ_2^2 ($p = 0.128$), indicating that in this analysis a history of renal A scores is adequately represented by a previous occurrence of a renal A.

The number of mucocutaneous A scores continues to have no significant effect on a patient's chance of presenting with active mucocutaneous disease; it leads to a change in deviance of 1.295 that is not significant on χ_2^2 ($p = 0.523$). The number of mucocutaneous B scores is significant when included in the model with time. However a quadratic factor is not necessary.

Time since a patient's first visit to the clinic has become significant on being included in the model with the number of mucocutaneous B scores; It leads to a change in deviance of 18.664 that is significant on on χ_1^2 ($p = 0.00002$). Both the time since the previous visit and the time since the patient was last observed to have scored a mucocutaneous A or B continue

197

to have no significant effect on a patient's chance of presenting with active mucocutaneous disease. Both together lead to a change in deviance of 0.100 that is not significant on χ_2^2 ($p = 0.951$).

Mucocutaneous damage, musculoskeletal damage and renal damage lead to changes in deviance of respectively 3.303 ($p = 0.069$), 1.350 ($p = 0.245$) and 0.184 ($p = 0.668$).

Variables that in this analysis did not significantly affect a patient's chance of presenting with active mucocutaneous disease were removed giving the results shown in table D.7.

Table D.7: Coefficients with adjusted odds ratios from the multivariate analysis (with only main effects) of active mucocutaneous disease vs. inactive mucocutaneous, musculoskeletal and renal disease.

	coefficient	odds ratio
A previous occurrence of a mucocutaneous B	0.894	2.446
Number of mucocutaneous B scores	0.187	1.206
A previous occurrence of a mucocutaneous A	1.106	3.021
A previous occurrence of a renal A	-1.098	0.334
A previous occurrence of a musculoskeletal A	-0.599	0.549
Time (years) since first clinic visit	-0.140	0.869

D.1.4 Interactions

See chapter 5.

D.2 The Renal Organ/System

D.2.1 Univariate analyses

Initial univariate analyses were carried out. The results are given with unadjusted odds ratios in table 5.7.

From table 5.7 it can be seen that a history of either renal A scores or renal B scores increases a patient's chance of presenting with active disease in the renal organ/system. In both cases both variables are significant indicating that the chance of presenting with active renal disease increases further with each additional observation of either a renal A or B.

Neither a history of musculoskeletal A scores nor a history of musculoskeletal B scores significantly affects a patient's chance of presenting with active renal disease. In both cases neither variable is significant.

A history of either mucocutaneous A scores or a history of mucocutaneous B scores significantly increases a patient's chance of presenting with active renal disease. Both variables are significant indicating that for both A and B scores the chance of presenting with active renal disease increases with each additional observation.

Time since the first clinic visit and time since the last clinic visit where the patient was observed to have active renal disease both significantly increase a patient's chance of presenting with active renal disease. The time since the previous clinic visit has no significant effect on a patient's chance of presenting with active renal disease.

Both renal and mucocutaneous damage significantly increase a patient's chance of presenting with active renal disease. Musculoskeletal damage does not have a significant effect.

History of disease activity in the organs/systems not chosen for detailed

analysis does not significantly affect a patient's chance of presenting at a clinic visit with active renal disease.

D.2.2   Analysis by individual Organ/System

Table D.8: Coefficients with adjusted odds ratios from the analysis by individual organ/system (renal) of active renal disease vs. inactive mucocutaneous, musculoskeletal and renal disease.

	coefficient	odds ratio
A previous occurrence of a renal B	1.535	4.643
Number of renal B scores	0.270	1.309
A previous occurrence of a renal A	3.931	50.980
Number of renal A scores	-4.244	
(Number of renal A scores)^2	1.122	$e^{-3.123+2.244ra*}$

* Where ra is the number of renal A's that the comparison patient had been observed to score at the previous visit.

The five variables representing a history of renal activity (the quadratic factor for the number of renal B's was found to be no longer necessary) as shown in table D.8 together lead to a change in deviance of 198.662 that is significant on χ_5^2 ($p \approx 0$). All variables significantly affect a patient's chance of presenting with active renal disease.

The four variables representing a history of musculoskeletal activity as shown in table D.9, together lead to a change in deviance of 4.655 that is not significant on χ_4^2 ($p = 0.325$). In this analysis neither a history of musculoskeletal A scores nor a history of musculoskeletal B scores significantly affects a patient's chance of presenting with active renal disease.

The four variables representing a history of mucocutaneous activity as

Table D.9: Coefficients with adjusted odds ratios from the analysis by individual organ/system (musculoskeletal) of active renal disease vs. inactive mucocutaneous, musculoskeletal and renal disease.

	coefficient	odds ratio
A previous occurrence of a musculoskeletal B	-0.027	0.974
Number of musculoskeletal B scores	-0.131	0.877
A previous occurrence of a musculoskeletal A	0.338	1.402
Number of musculoskeletal A scores	0.076	1.079

Table D.10: Coefficients with adjusted odds ratios from the analysis by individual organ/system (mucocutaneous) of active renal disease vs. inactive mucocutaneous, musculoskeletal and renal disease.

	coefficient	odds ratio
A previous occurrence of a mucocutaneous B	0.209	1.232
Number of mucocutaneous B scores	0.045	1.046
A previous occurrence of a mucocutaneous A	0.381	1.464
Number of mucocutaneous A scores	0.075	1.078

shown in table D.10 together lead to a change in deviance of 8.501 that is not significant on χ^2_4 ($p = 0.075$). A history of mucocutaneous A scores has lost significance. Both variables together lead to a change in deviance of 3.373 that is not significant on χ^2_2 ($p = 0.185$). This is possibly because all patients with a history of mucocutaneous A scores have a history of mucocutaneous B scores as shown in table D.11.

The number of mucocutaneous B scores has also lost significance; it leads to a change in deviance of 0.907 that is not significant on χ^2_1 ($p = 0.341$).

Table D.11: Numbers of patients with histories of mucocutaneous A and B scores.

	No history of mucocutaneous A scores	History of mucocutaneous A scores
No history of mucocutaneous B scores	133	0
History of a mucocutaneous B	165	65

This indicates that in this analysis a previous occurrence of a mucocutaneous B adequately represents a history of mucocutaneous activity.

D.2.3 Multivariate Analysis

All variables were now included in a single model with only main effects. A history of disease activity in the five organs/systems not chosen for detailed analysis continues not to significantly affect a patient's chance of presenting with active renal disease. It leads to a change in deviance of 0.836 that is not significant on χ_1^2 ($p = 0.361$).

Time since the first clinic visit and time since the patient was last observed to score a renal A or B, both lose significance when included in the model with the number of times a patient has scored a renal B or the number of times a patient has scored a renal A. This indicates that in the univariate analysis these variables were picking up the effect of the number of renal B's or A's scored. All three time variables together lead to a change in deviance of 3.168 that is not significant on χ_3^2 ($p = 0.366$).

A history of mucocutaneous disease activity loses all significance once a history of renal B scores has been adjusted for. All 4 variables representing a history of mucocutaneous disease activity together lead to a change in

Table D.12: The % of visits with details of patients' histories of mucocutaneous A, mucocutaneous B and renal B scores.

No history of mucocutaneous B scores	$\frac{66}{1835} = 3.60\%$
History of mucocutaneous B scores	$\frac{58}{1120} = 5.18\%$

	No history of renal B scores	History of renal B scores
No history of mucocutaneous B scores	$\frac{26}{1545} = 1.68\%$	$\frac{40}{290} = 13.79\%$
History of mucocutaneous B scores	$\frac{8}{796} = 1.00\%$	$\frac{50}{324} = 15.43\%$

No history of mucocutaneous A scores	$\frac{104}{2668} = 3.90\%$
History of mucocutaneous A scores	$\frac{20}{287} = 6.97\%$

	No history of renal B scores	History of renal B scores
No history of mucocutaneous A scores	$\frac{33}{2203} = 1.04\%$	$\frac{71}{465} = 15.27\%$
History of mucocutaneous A scores	$\frac{1}{138} = 0.72\%$	$\frac{19}{149} = 12.75\%$

deviance of 1.033 that is not significant on χ_4^2 ($p = 0.905$).

The % of visits where a patient presents with active renal disease are given with details of mucocutaneous A and B and renal B scores in table D.12.

A history of musculoskeletal A scores continues in this analysis to have no significant effect on a patient's chance of presenting with active renal disease. Both variables together lead to a change in deviance of 0.672 that is not significant on χ_2^2 ($p = 0.715$).

A previous occurrence of a musculoskeletal B continues not to significantly affect a patient's chance of presenting with active renal disease. It leads to a

Table D.13: The % of visits with details of patients' histories of musculoskeletal B and renal B scores.

Number of musculoskeletal B scores					
0	1	2	3	4	≥ 5
$\frac{66}{1425} = 4.63\%$	$\frac{32}{715} = 4.48\%$	$\frac{9}{360} = 2.50\%$	$\frac{9}{212} = 4.24\%$	$\frac{5}{96} = 5.21\%$	$\frac{3}{144} = 2.08\%$

	Number of musculoskeletal B scores		
	0	1	2
No history of renal B scores	$\frac{25}{1172} = 2.13\%$	$\frac{4}{565} = 0.71\%$	$\frac{3}{300} = 1.00\%$
History of renal B scores	$\frac{41}{253} = 16.21\%$	$\frac{28}{150} = 18.67\%$	$\frac{6}{60} = 10.00\%$

	Number of musculoskeletal B scores		
	3	4	≥ 5
No history of renal B scores	$\frac{0}{150} = 0\%$	$\frac{2}{66} = 3.03\%$	$\frac{0}{88} = 0\%$
History of renal B scores	$\frac{9}{62} = 14.52\%$	$\frac{3}{30} = 10.00\%$	$\frac{3}{56} = 5.36\%$

change in deviance of 1.363 that is not significant on χ_1^2 ($p = 0.243$). However the number of times a patient has been observed to score a musculoskeletal B has become significant when included in the model with a previous occurrence of a renal B; it decreases a patient's chance of presenting with active renal disease. It is clear from table D.13 that gives the % of visits where a patient presents with active renal disease, that the effect of the number of times the patient has scored a musculoskeletal B becomes more evident once a history of renal B scores has been adjusted for.

The number of renal A's lost significance once the number of musculoskeletal B's was adjusted for. It leads to a change in deviance of 2.798 that is not significant on χ_2^2 ($p = 0.247$).

Renal damage, mucocutaneous damage and musculoskeletal damage lead

Table D.14: Coefficients with adjusted odds ratios from the multivariate analysis (with only main effects) of active renal disease vs. inactive mucocutaneous, musculoskeletal and renal disease.

	coefficient	odds ratio	p value
Previous occurrence of a renal B	1.578	4.846	≈ 0
Number of renal B scores	0.272	1.312	0.002
Previous occurrence of a renal A	0.732	2.032	0.004
Number of musculoskeletal B scores	-0.244	0.709	0.001
Renal damage	1.247	3.478	0.00004

to changes in deviance of respectively 19.867 ($p = 0.001$), 1.992 ($p = 0.158$) and 0.064 ($p = 0.800$).

Variables that did not significantly affect a patient's chance of presenting with active renal disease were removed giving the results shown with adjusted odds ratios in table D.14.

D.2.4 Interactions

See Chapter 5.

D.3 The Musculoskeletal Organ/System

D.3.1 Univariate analysis

Initial univariate analyses were carried out. The results with unadjusted odds ratios are given in table 5.13.

A history of either musculoskeletal A or musculoskeletal B scores increases a patient's chance of presenting with active disease in the musculoskeletal

organ/system. For both A and B scores both variables are significant. The effect of the number of times a patient has scored a musculoskeletal B is not linear and a quadratic factor has been included, indicating in this case that the effect of each additional observation decreases as the number of observations increases as shown in table D.15.

Table D.15: Unadjusted odds ratios for each additional musculoskeletal B score.

Number of musculoskeletal B's scored by the previous visit	Odds ratios for a musculoskeletal B score
0	1.582
1	1.481
2	1.387
3	1.298
4	1.215
5	1.138
6	1.065
7	0.997
8	0.933
9	0.874

A history of renal A scores does not significantly affect a patient's chance of presenting with active musculoskeletal disease, neither variable is significant.

A history of renal B scores however significantly decreases a patient's chance of presenting with active musculoskeletal disease. A previous occurrence of a renal B is not significant, and the number of times a patient has scored a renal B is not linear and a quadratic factor was included. This indicates that the effect of each observation changes with each additional

observation.

A history of mucocutaneous A scores does not significantly affect a patient's chance of presenting with active musculoskeletal disease. A history of mucocutaneous B scores increases a patient's chance of presenting with active musculoskeletal disease. A previous occurrence of a mucocutaneous B is significant whereas the number of mucocutaneous B's is not.

A patient's chance of presenting with active musculoskeletal disease decreases with the time since the last clinic visit at that the patient was observed to have active musculoskeletal disease.

Time since the previous clinic visit and the time since the first clinic visit do not significantly affect a patient's chance of presenting with active musculoskeletal disease.

Musculoskeletal damage significantly increases a patient's chance of presenting with active musculoskeletal disease. Neither mucocutaneous damage nor renal damage have any significant effect.

History of disease activity in the organs/systems not chosen for detailed analysis does not affect a patient's chance of presenting at a clinic visit with active musculoskeletal disease.

D.3.2 Analysis by individual organ/system

The five variables representing a history of musculoskeletal activity shown in table D.16, together lead to a change in deviance of 112.264 that is significant on χ_5^2 ($p \approx 0$). However a history of musculoskeletal A scores has lost significance. It leads to a change in deviance of 4.827 that is not significant on χ_2^2 ($p = 0.090$). This is possibly due to the fact that most patients with a history of musculoskeletal A scores have a history of musculoskeletal B scores as shown in table D.17.

Table D.16: Coefficients with adjusted odds ratios from the analysis by individual organ/system (musculoskeletal) of active musculoskeletal disease vs. inactive mucocutaneous, musculoskeletal and renal disease

	coefficient	odds ratio
A previous occurrence of a musculoskeletal B	-0.033	0.963
Number of musculoskeletal B scores	0.478	
(Number of musculoskeletal B scores)^2	-0.034	$e^{0.444-0.068m_{u}b}$
A previous occurrence of a musculoskeletal A	-0.047	0.954
Number of musculoskeletal A scores	0.204	1.226

Table D.17: Coefficients with adjusted odds ratios from the analysis by individual organ/system (musculoskeletal) of active musculoskeletal disease vs. inactive mucocutaneous, musculoskeletal and renal disease

	No history of musculoskeletal A scores	History of musculoskeletal A scores
No history of musculoskeletal B scores	105	5
History of musculoskeletal B scores	206	51

A previous occurrence of a musculoskeletal B has lost significance on being included in a model with the number of times a patient has scored a musculoskeletal B. It leads to a change in deviance of 0.089 that is not significant on χ^2_1 ($p = 0.765$). A history of musculoskeletal disease activity is therefore in this analysis, adequately represented by the number of times a patient has scored a musculoskeletal B.

The five variables representing a history of renal activity as shown in table D.18, together lead to a change in deviance of 14.411 that is significant

on χ_5^2 $(p = 0.013)$.

Table D.18: Coefficients with adjusted odds ratios from the analysis by individual organ/system (renal) of active musculoskeletal disease vs. inactive mucocutaneous, musculoskeletal and renal disease

	coefficient	odds ratio
A previous occurrence of a renal B	-0.369	0.691
Number of renal B scores	-0.059	
(Number of renal B scores)^2	0.056	$e^{-0.003+0.112rb}$
A previous occurrence of a renal A	-0.467	0.608
Number of renal A scores	-0.071	0.931

A history of renal A scores leads to a change in deviance of 4.072 that is not significant on χ_2^2 $(p = 0.131)$, and the quadratic factor appears no longer to be necessary to describe the number of renal B scores. The number of times the patient has scored a renal B, and a previous occurrence of a renal B together lead to a change in deviance of 9.365 that is significant on χ_2^2 $(p = 0.001)$. Both variables shown in table D.19 significantly add to the model.

Table D.19: Coefficients with adjusted odds ratios from the analysis by individual organ/system (renal) of active musculoskeletal disease vs. inactive mucocutaneous, musculoskeletal and renal disease

	coefficient	odds ratio
A previous occurrence of a renal B	-0.764	0.466
Number of renal B scores	0.226	1.254

Table D.20: Coefficients with adjusted odds ratios from the analysis by individual organ/system (mucocutaneous) of active musculoskeletal disease vs. inactive mucocutaneous, musculoskeletal and renal disease

	coefficient	odds ratio
A previous occurrence of a mucocutaneous B	0.240	1.272
Number of mucocutaneous B scores	0.006	1.006
A previous occurrence of a mucocutaneous A	0.021	1.022
Number of mucocutaneous A scores	-0.574	
(Number of mucocutaneous A scores)^2	0.191	$e^{-0.384+0.382ma}$

The five variables shown in table D.20 together lead to a change in deviance of 10.775 that is not significant on χ_5^2 ($p = 0.056$).

D.3.3 Multivariate Analysis

All variables were now included in a single model with only main effects.

A history of renal A scores continues to have no significant effect on a patient's chance of presenting with active musculoskeletal disease. It leads to a change in deviance of 1.895 that is not significant on χ_2^2 ($p = 0.388$).

A history of mucocutaneous activity is not significant It leads to a change in deviance of 9.264 that is not significant on χ_5^2 ($p = 0.099$). (A previous occurrence of a mucocutaneous B leads to a change in deviance of 0.879 that is not significant on χ_1^2 ($p = 0.348$), and the number of mucocutaneous A scores leads to a change in deviance of 3.262 that is not significant on χ_2^2 ($p = 0.196$).) This is possibly because as has been previously noted, most patients who have a history of mucocutaneous B scores also have a history of musculoskeletal B scores, as is shown in table D.21.

A history of renal B scores continues as was indicated by the previous

210

Table D.21: The number of patients with histories of mucocutaneous B and musculoskeletal B scores

	No history of mucocutaneous B scores	History of mucocutaneous B scores
No history of musculoskeletal B scores	51	53
History of musculoskeletal B scores	82	175

analyses to significantly decrease a patient's chance of presenting with active musculoskeletal disease. In this analysis however it appears that having adjusted for a history of musculoskeletal B scores a history of renal B scores is adequately represented by a previous occurrence of a renal B. The number renal B scores leads to a change in deviance of 1.003 that is not significant on χ_2^2 ($p = 0.606$).

A history of musculoskeletal A scores has remained significant in this analysis. Both variables together lead to a change in deviance of 7.595 that is significant on χ_2^2 ($p = 0.022$). However a previous occurrence of a musculoskeletal A leads to a change in deviance of 0.012 that is not significant on χ_1^2 ($p = 0.913$) and the number of times a patient has scored a musculoskeletal A leads to a change in deviance of 3.098 that is not significant on χ_1^2 ($p = 0.078$). For simplicity a previous occurrence of a musculoskeletal A will be used in this analysis.

A previous occurrence of a musculoskeletal B has again lost significance on being included in a model with the number of musculoskeletal B's; it leads to a change in deviance of 0.058 that is not significant on χ_1^2 ($p = 0.810$).

Time since the first clinic visit has become significant on being included in the model with a history of musculoskeletal B scores; it leads to a change

in deviance of 4.825 that is significant on χ_1^2 ($p = 0.028$).

Time since the previous clinic visit has become significant on being included in the model with a history of musculoskeletal B scores; it leads to a change in deviance of 5.480 that is significant on χ_1^2 ($p = 0.019$).

Time since the patient was last observed to score a musculoskeletal A or B continues to significantly affect a patient's chance of presenting with active musculoskeletal disease. It leads to a change in deviance of 10.671 that is significant on χ_1^2 ($p = 0.001$).

Musculoskeletal damage, mucocutaneous damage and renal damage, lead to changes in deviance of respectively 5.160 ($p = 0.023$), 0.091 ($p = 0.764$) and 3.341 ($p = 0.068$). Although musculoskeletal damage has a significant effect on a patient's chance of presenting with active musculoskeletal disease, including it in the model causes a previous occurrence of a musculoskeletal A to lose significance. This indicates that either could be included in the model, but that there is no advantage in including both. It has been decided to use a previous occurrence of a musculoskeletal A.

All other variables that were significant in the univariate analyses continue to significantly affect a patient's chance of presenting with active musculoskeletal disease.

Variables that did not significantly affect a patient's chance of presenting with active musculoskeletal disease were removed giving the results given in table D.22.

D.3.4 Interactions

See Chapter 5.

Table D.22: Coefficients with adjusted odds ratios from the multivariate analysis (with only main effects) of active musculoskeletal disease vs. inactive mucocutaneous, musculoskeletal and renal disease.

	coefficient	odds ratio	p value
Number of musculoskeletal B scores	0.695		
(Number of musculoskeletal B scores)^2	-0.050	$e^{0.645-0.100m_ub}$	≈ 0
A previous occurrence of a musculoskeletal A	0.341	1.406	0.032
A previous occurrence of a renal B	-0.523	0.593	0.002
History of disease activity in all other organs/systems	-0.419	0.658	0.003
Time (years) since first clinic visit	-0.066	0.936	0.032
Time (years) since previous clinic visit	0.403	1.497	0.010
Time (years) since last visit with active musculoskeletal disease	-0.172	0.842	0.002

D.4 Simultaneous Mucocutaneous and Musculoskeletal disease

D.4.1 Univariate analyses

Initial univariate analyses were carried out. The results are given in table 5.21.

A history of mucocutaneous B scores significantly increases a patient's chance of presenting with active disease in both the mucocutaneous and musculoskeletal systems. Both variables are significant. A history of mucocutaneous A scores does not significantly affect a patient's presenting with active disease in both the mucocutaneous and musculoskeletal systems.

A history of musculoskeletal B scores significantly increases a patient's chance of presenting with active disease in both the mucocutaneous and

musculoskeletal systems. Both variables are significant. A history of musculoskeletal A scores does not significantly affect a patient's presenting with active disease in both the mucocutaneous and musculoskeletal systems.

A history of renal B scores does not significantly affect a patient's chance of presenting with active disease in the mucocutaneous and musculoskeletal systems.

No patients with a history of renal A scores have presented at a clinic visit with active mucocutaneous and musculoskeletal disease. The numbers of patients are given in table D.23. As a result of this no further mention of a history of renal A scores will be made in this section.

Table D.23: The numbers of patients with a history of simultaneous active mucocutaneous and musculoskeletal disease and details of their history of renal A scores.

	No history of mucocutaneous and musculoskeletal disease	History of mucocutaneous and musculoskeletal disease
No history of renal A scores	300	12
History of renal A scores	22	0

A patient's chance of presenting with active mucocutaneous and musculoskeletal disease decreases with the time since their first clinic visit and with the time since they were last observed to score a musculoskeletal A or B.

The time since the previous visit and the time since they were last observed to score a mucocutaneous A or B have no significant effect on a patient's chance of presenting with active mucocutaneous and musculoskeletal disease.

Musculoskeletal damage significantly affects a patient's chance of present-

ing with both active mucocutaneous and musculoskeletal disease. Neither mucocutaneous nor renal damage have a significant effect.

A history of disease activity in the five organs/systems not chosen for detailed analysis does not significantly affect a patient's chance of presenting with active disease in both the mucocutaneous and musculoskeletal organs/systems.

D.4.2 Analysis by individual Organ/System

Table D.24: Coefficients with adjusted odds ratios from the analysis by individual organ/system (musculoskeletal) of active musculoskeletal and mucocutaneous disease vs. inactive mucocutaneous, musculoskeletal and renal disease.

	coefficient	odds ratio
A previous occurrence of a musculoskeletal B	0.419	1.521
Number of musculoskeletal B scores	0.134	1.143
A previous occurrence of a musculoskeletal A	-0.568	0.567
Number of musculoskeletal A scores	0.098	1.103

The four variables representing a history of active musculoskeletal disease, as shown in table D.24, together lead to a change in deviance of 10.8068 that is significant on χ_4^2 ($p = 0.031$). A history of musculoskeletal A scores continues to have no significant effect on a patient's chance of presenting with active mucocutaneous and musculoskeletal disease; it leads to a change in deviance of 1.2066 that is not significant on χ_2^2 ($p = 0.547$). A previous occurrence of a musculoskeletal B loses significance on being included in the model with the number of times a patient has scored a musculoskeletal B; it leads to a change

in deviance of 1.5562 that is not significant on χ_1^2 ($p = 0.212$). This indicates that in this analysis a history of active musculoskeletal disease is adequately represented by the number of times a patient has scored a musculoskeletal B.

Table D.25: Coefficients with adjusted odds ratios from the analysis by individual organ/system (mucocutaneous) of active musculoskeletal and mucocutaneous disease vs. inactive mucocutaneous, musculoskeletal and renal disease.

	coefficient	odds ratio
A previous occurrence of a mucocutaneous B	0.557	1.746
Number of times observed to score a mucocutaneous B	0.079	1.082
History of mucocutaneous A activity	0.481	1.618
Number of times observed to score a mucocutaneous A	-0.375	0.687

The four variables representing a history of mucocutaneous activity as shown in table D.25, together lead to a change in deviance of 9.4332 that is not significant on χ_4^2 ($p = 0.051$). A history of mucocutaneous A scores continues to have no significant effect on a patient's chance of presenting with active mucocutaneous and musculoskeletal disease; it leads to a change in deviance of 0.8876 that is not significant on χ_2^2 ($p = 0.642$).

Neither a previous occurrence of a mucocutaneous B score nor the number of mucocutaneous B scores now adds significantly to the model. This indicates that either variable adequately represents a history of mucocutaneous B scores.

The two variables representing a history of renal activity as shown in table D.26, together lead to a change in deviance of 1.8068 that is not significant on χ_2^2 ($p = 0.405$).

Table D.26: Coefficients with adjusted odds ratios from the analysis by individual organ/system (renal) of active vs. not active musculoskeletal and mucocutaneous disease.

	coefficient	odds ratio
A previous occurrence of a renal B	-0.104	0.901
Number of renal B scores	-0.184	0.832

D.4.3 Multivariate Analysis

All variables were now included in a single model with only main effects.

A history of disease activity in the five organs/systems not chosen for detailed analysis continues to appear not to significantly affect a patient's chance of presenting with both active mucocutaneous and musculoskeletal disease; it leads to a change in deviance of 2.842 that is not significant on χ_1^2 $(p = 0.093)$.

As in the previous analysis a history of active musculoskeletal disease is adequately represented by the number of times the patient has scored a musculoskeletal B. The three other variables representing a history of active musculoskeletal disease together lead to a change in deviance of 1.9989 that is not significant on χ_3^2 $(p = 0.573)$.

A history of mucocutaneous A scores continues not to significantly affect a patient's chance of presenting with both active mucocutaneous and musculoskeletal disease; it leads to a change in deviance of 0.2893 that is not significant on χ_2^2 $(p = 0.865)$. As in the previous analysis it appears that either the binary variable representing a history of mucocutaneous B scores or the number of mucocutaneous B scores adequately describes a history of mucocutaneous B scores. For simplicity the binary variable will be used.

The number of times the patient has scored a mucocutaneous B leads to a change in deviance of 0.9053 that is not significant on χ_1^2 ($p = 0.341$).

A history of renal B scores does not significantly affect a patient's chance of presenting with both active mucocutaneous and musculoskeletal disease; it leads to a change in deviance of 0.0978 that is not significant on χ_2^2 ($p = 0.952$).

Time since the first clinic visit has become significant on being included in the model with the number of musculoskeletal B scores; it leads to a change in deviance of 25.339 that is significant on χ_1^2 ($p \approx 0$).

Time since the last visit with active musculoskeletal disease has lost significance on being included in the model with the time since the first clinic visit; it leads to a change in deviance of 2.209 that is not significant on χ_1^2 ($p = 0.137$).

Both time since the patient was last observed to score a mucocutaneous A or B, and the time since the previous visit, continue to have no significant effect on a patient's chance of presenting with active mucocutaneous and musculoskeletal disease. Both variables together lead to a change in deviance of 0.512 that is not significant on χ_2^2 ($p = 0.474$).

Musculoskeletal damage, mucocutaneous damage and renal damage, lead to changes in deviance of respectively 5.433 ($p = 0.020$), 0.412 ($p = 0.521$) and 0.371 ($p = 0.543$).

The variables that did not have a significant effect on a patient's chance of presenting with active mucocutaneous and musculoskeletal disease were removed, giving the results shown in table D.27

(Musculoskeletal damage loses significance when the model is fitted using generalized estimating equations).

Table D.27: Coefficients with adjusted odds ratios from the multivariate analysis of active musculoskeletal and mucocutaneous disease vs. inactive mucocutaneous, musculoskeletal and renal disease. (Not fitted using generalized estimating equations).

	coefficient	odds ratio
A previous occurrence of a mucocutaneous B	1.003	2.725
Number of musculoskeletal B scores	0.271	1.311
Musculoskeletal damage	0.806	2.238
time (years) since the first clinic visit	-0.369	0.691

D.4.4 Interactions

No significant interactions were found.

D.5 Active renal disease with active disease in one or both of the musculoskeletal and mucocutaneous organs/systems.

In this analysis a patient presenting with renal disease should be understood as a patient presenting with both active renal disease and active mucocutaneous disease, or both active renal disease and active musculoskeletal disease or active disease in all three organs/systems.

D.5.1 Univariate analyses

Initial univariate analyses were carried out. The results are given with unadjusted odds ratios in table 5.23.

A history of renal B scores significantly increases a patient's chance of presenting with active renal disease Both variables are significant. The number of renal B's is not linear and a quadratic factor has been added. This indicates that the effect decreases with each subsequent renal B score.

A history of renal A scores increases a patient's chance of presenting with active renal disease. The number of renal A scores is significant indicating that a patient's chance of presenting with active renal disease increases with each subsequent observation of a renal A. (A previous occurrence of a renal A is not significant).

A history of mucocutaneous B scores significantly increases a patient's chance of presenting with active renal disease. Both variables are significant indicating that a patient's chance of presenting with active renal disease increases with each subsequent observation of a mucocutaneous B. A history of mucocutaneous A scores also significantly affects a patient's chance of presenting with active renal disease. The number of mucocutaneous A scores is significant, again indicating that the patient's chance of presenting with active disease increases with each subsequent observation of a mucocutaneous A.

A history of disease activity in the musculoskeletal organ/system does not significantly affect a patient's chance of presenting with active renal disease.

None of renal, mucocutaneous or musculoskeletal damage affect a patient's chance of presenting with active renal disease.

A history of disease activity in the organs/systems not chosen for detailed analysis does not significantly affect a patient's chance of presenting with active renal disease.

The times since the fist clinic visit, the previous visit, the last observation of a renal A or B, a mucocutaneous A or B, or a musculoskeletal A or B have

no significant effect on a patient's chance of presenting with active renal disease.

D.5.2 Analysis by individual Organ/System

The five variables representing a history of renal activity as shown in table D.28 together lead to a change in deviance of 28.6489 that is significant on χ_5^2 ($p = 0.00003$).

A history of renal A scores leads to a change in deviance of 4.7684 that is not significant on χ_2^2 ($p = 0.092$).

Table D.28: Coefficients with adjusted odds ratios from the analysis by individual organ/system (renal) of active renal disease vs. inactive mucocutaneous, musculoskeletal and renal disease.

	coefficient	odds ratio
A previous occurrence of a renal B	0.832	2.299
Number of renal B scores	1.037	
(Number of renal B scores)^2	-0.151	$e^{0.886-0.302rb}$
History of renal A activity	-2.622	0.073
Number of times observed to score a renal A	986	2.680

The number of times a patient has scored a renal B leads to a change in deviance of 1.6638 that is not significant on χ_2^2 ($p = 0.435$). This indicates that in this analysis a history of active renal disease is adequately represented by the binary variable indicating whether a patient has had a previous occurrence of a renal B.

The four variables representing a history of active musculoskeletal disease, as shown in table D.29 together lead to a change in deviance of 4.1636

221

Table D.29: Coefficients with adjusted odds ratios from the analysis by individual organ/system (musculoskeletal) of active renal disease vs. inactive mucocutaneous, musculoskeletal and renal disease.

	coefficient	odds ratio
A previous occurrence of a musculoskeletal B	0.292	1.339
Number of musculoskeletal B scores	0.027	1.027
A previous occurrence of a musculoskeletal A	-1.022	0.360
Number of musculoskeletal A scores	0.674	1.962

that is not significant on χ_4^2 ($p = 0.384$). This indicates that in this analysis a history of active musculoskeletal disease does not significantly affect a patient's chance of presenting with active renal disease.

The four variables representing a history of mucocutaneous activity as shown in table D.30 8.896 that is not significant on χ_4^2 ($p = 0.064$). However

Table D.30: Coefficients with adjusted odds ratios from the analysis by individual organ/system (mucocutaneous) of active renal disease vs. inactive mucocutaneous, musculoskeletal and renal disease.

	coefficient	odds ratio
A previous occurrence of a mucocutaneous B	0.398	1.489
Number of times observed to score a mucocutaneous B	0.119	1.126
History of mucocutaneous A activity	-0.273	0.761
Number of times observed to score a mucocutaneous A	0.390	1.478

a history of mucocutaneous A scores has lost significance, it leads to a change in deviance of 2.1603 that is not significant on χ_2^2 ($p = 0.340$). Neither the number of mucocutaneous B scores nor a previous occurrence of a mucocutaneous B adds significantly to the model once the other is included, indicating

that in this analysis either variable adequately represents a history of muco-cutaneous activity.

D.5.3 Multivariate Analysis

All variables were now included in a single model with only main effects.

A history of disease activity in the five organs/systems not chosen for detailed analysis continues to appear not to significantly affect a patient's chance of presenting with active renal disease. It leads to a change in deviance of 0.9639 that is not significant on χ_1^2 ($p = 0.326$).

A history of musculoskeletal disease activity continues to have no significant effect on a patient's chance of presenting with active renal disease. It leads to a change in deviance of 2.2563 that is not significant on χ_4^2 ($p = 0.669$).

A history of mucocutaneous B scores has lost significance on being included in the model with a previous occurrence of a renal B. A history of mucocutaneous disease activity leads to a change in deviance of 6.645 that is not significant on χ_4^2 ($p = 0.156$).

A history of renal A scores continues not to significantly affect a patient's chance of presenting with active renal disease; it leads to a change in deviance of 4.7451 that is not significant on χ_2^2 ($p = 0.093$).

The time since the first clinic visit continues to have no significant effect on a patient's chance of presenting with active renal disease; It leads to a change in deviance of 0.006 that is not significant on χ_1^2 ($p = 0.938$).

The times since the patient was last observed to score a mucocutaneous or musculoskeletal A or B continue to have no significant effect on a patient's chance of presenting with active renal disease. Both variables together lead to a change in deviance of 0. 081 that is not significant on χ_2^2 ($p = 0.960$).

Table D.31: Coefficients with adjusted odds ratios from the multivariate analysis of active renal disease vs. inactive mucocutaneous, musculoskeletal and renal disease.

	coefficient	odds ratio
A previous occurrence of a renal B	2.955	19.206
Time (years) since previous clinic visit	0.704	2.021
Time (years) since last scored a renal A or B	-0.963	0.382

However the time since the patient was last observed to score a renal A or B and the time since the previous visit have become significant on being included in the model with a previous occurrence of a renal B. The time since the previous visit leads to a change in deviance of 4.221 that is significant on χ_1^2 ($p = 0.040$). The time since the last observed renal A or B leads to a change in deviance of 10.181 that is significant on χ_1^2 ($p = 0.001$).

Finally The number of renal B scores leads to a change in deviance of 0.7318 that is not significant on χ_1^2 ($p = 0.392$).

Renal damage, mucocutaneous damage and musculoskeletal damage, lead to changes in deviance of respectively 0.058 ($p = 0.809$), 0.396 ($p = 0.521$) and 0.236 ($p = 0.627$).

Variables that did not significantly affect a patient's chance of presenting with active disease were removed giving the result given in table D.31.

(Time since the patient was last observed to score a renal A or B loses significance when the model was fit using generalized estimating equations).

D.5.4 Interactions

No significant interactions were found.

Appendix E

The relative efficiency of the logistic model to the exponential when observations are made at three distinct time points

Probability of failing before $t_1 = 10\%$

Probability of failing before t_2	Probability of failing before t_3								
	10%	20%	30%	40%	50%	60%	70%	80%	90%
10%	1	0.961	0.873	0.759	0.631	0.497	0.363	0.236	0.124
20%		0.968	0.916	0.828	0.716	0.589	0.453	0.316	0.187
30%			0.901	0.843	0.754	0.642	0.514	0.377	0.241
40%				0.815	0.751	0.659	0.544	0.415	0.278
50%					0.712	0.641	0.543	0.426	0.295
60%						0.590	0.512	0.409	0.289
70%							0.451	0.365	0.258
80%								0.295	0.205
90%									0.134

Probability of failing before $t_1 = 20\%$

Probability of failing before t_2	Probability of failing before t_3								
	10%	20%	30%	40%	50%	60%	70%	80%	90%
10%	0.961	0.968	0.916	0.828	0.716	0.589	0.483	0.316	0.187
20%		1	0.974	0.907	0.809	0.669	0.552	0.408	0.264
30%			0.977	0.935	0.858	0.752	0.624	0.480	0.330
40%				0.919	0.865	0.778	0.664	0.528	0.378
50%					0.834	0.768	0.672	0.548	0.403
60%						0.723	0.645	0.537	0.403
70%							0.586	0.494	0.373
80%								0.419	0.314
90%									0.227

Probability of failing before $t_1 = 30\%$

Probability of failing before t_2	Probability of failing before t_3								
	10%	20%	30%	40%	50%	60%	70%	80%	90%
10%	0.873	0.916	0.901	0.844	0.754	0.642	0.514	0.377	0.240
20%		0.974	0.977	0.935	0.858	0.752	0.624	0.480	0.330
30%			1	0.979	0.920	0.828	0.708	0.565	0.408
40%				0.980	0.942	0.867	0.761	0.626	0.469
50%					0.925	0.870	0.781	0.658	0.507
60%						0.836	0.765	0.658	0.516
70%							0.713	0.622	0.493
80%								0.548	0.434
90%									0.336

Probability of failing before $t_1 = 40\%$

Probability of failing before t_2	Probability of failing before t_3								
	10%	20%	30%	40%	50%	60%	70%	80%	90%
10%	0.759	0.828	0.844	0.815	0.751	0.659	0.544	0.415	0.278
20%		0.907	0.935	0.919	0.865	0.778	0.664	0.528	0.378
30%			0.979	0.980	0.942	0.867	0.761	0.626	0.469
40%				1	0.980	0.923	0.829	0.701	0.544
50%					0.981	0.942	0.865	0.749	0.598
60%						0.924	0.865	0.764	0.622
70%							0.825	0.741	0.612
80%								0.675	0.559
90%									0.457

Probability of failing before $t_1 = 50\%$

Probability of failing before t_2	Probability of failing before t_3								
	10%	20%	30%	40%	50%	60%	70%	80%	90%
10%	0.631	0.716	0.754	0.751	0.712	0.641	0.543	0.426	0.295
20%		0.809	0.858	0.865	0.834	0.768	0.672	0.548	0.403
30%			0.920	0.942	0.925	0.870	0.781	0.658	0.507
40%				0.980	0.981	0.942	0.865	0.749	0.598
50%					1	0.980	0.919	0.816	0.669
60%						0.980	0.938	0.850	0.713
70%							0.916	0.845	0.721
80%								0.793	0.683
90%									0.586

Probability of failing before $t_1 = 60\%$

Probability of failing before t_2	Probability of failing before t_3								
	10%	20%	30%	40%	50%	60%	70%	80%	90%
10%	0.497	0.589	0.642	0.659	0.641	0.590	0.512	0.409	0.289
20%		0.689	0.752	0.778	0.768	0.723	0.645	0.537	0.403
30%			0.828	0.867	0.870	0.836	0.765	0.658	0.516
40%				0.923	0.942	0.924	0.865	0.764	0.622
50%					0.980	0.980	0.938	0.850	0.713
60%						1	0.978	0.907	0.780
70%							0.977	0.926	0.813
80%								0.895	0.799
90%									0.717

Probability of failing before $t_1 = 70\%$

Probability of failing before t_2	Probability of failing before t_3								
	10%	20%	30%	40%	50%	60%	70%	80%	90%
10%	0.363	0.453	0.514	0.544	0.543	0.512	0.451	0.365	0.258
20%		0.552	0.624	0.664	0.672	0.645	0.586	0.494	0.373
30%			0.708	0.761	0.781	0.765	0.713	0.622	0.493
40%				0.829	0.865	0.865	0.825	0.741	0.612
50%					0.919	0.938	0.916	0.845	0.721
60%						0.978	0.977	0.926	0.813
70%							1	0.972	0.877
80%								0.970	0.896
90%									0.843

Probability of failing before $t_1 = 80\%$

Probability of failing before t_2	Probability of failing before t_3								
	10%	20%	30%	40%	50%	60%	70%	80%	90%
10%	0.236	0.316	0.377	0.415	0.426	0.409	0.365	0.295	0.205
20%		0.408	0.480	0.528	0.548	0.537	0.494	0.419	0.314
30%			0.565	0.626	0.658	0.658	0.622	0.548	0.434
40%				0.701	0.749	0.764	0.741	0.675	0.559
50%					0.816	0.850	0.845	0.793	0.683
60%						0.907	0.926	0.895	0.799
70%							0.972	0.970	0.896
80%								1	0.957
90%									0.949

Probability of failing before $t_1 = 90\%$

Probability of failing before t_2	Probability of failing before t_3								
	10%	20%	30%	40%	50%	60%	70%	80%	90%
10%	0.124	0.187	0.241	0.278	0.295	0.289	0.258	0.205	0.134
20%		0.264	0.330	0.378	0.403	0.403	0.373	0.314	0.227
30%			0.408	0.469	0.507	0.516	0.493	0.434	0.336
40%				0.544	0.598	0.622	0.612	0.559	0.457
50%					0.669	0.713	0.721	0.683	0.586
60%						0.780	0.813	0.799	0.717
70%							0.877	0.896	0.843
80%								0.957	0.949
90%									1

Bibliography

Aalen, O., Fosen, J., Fekjaer, H., Borgan, O., and Husebye, E. (2003). Dynamic analysis of multivariate failure time data.

Allen, E.J. and Farewell, V.T. (2002). *Appendix to:Sentence conjunctions in the gospel of Matthew*. Sheffield Academic Press.

Barlow, J. E., Austin, H. A., and Tsokos, G. C. (1987). Lupus nephritis. *Annals of Internal Medicine*, **106**, 79–94.

Begg and Grey (1984). Calculation of polychotomous logistic regression parameters using individualized regressions. *Biometrika*, **71**, 11–8.

Belmont, H. Lupus clinical overview.

Bombardier, C., Gladman, D., Urowitz, M.B., *et al.* (1992). The development and validation of the SLE disease activity index (SLEDAI). *Arthritis and Rheumatism*, **35**, 630–40.

Boumpas, D. T., J., B., and Austin, H. A. (1995). Systemic lupus erythematosus: emerging concepts. part 2: Dermatologic and joint disease, the antiphospholipid antibody syndrome, pregnancy and hormonal therapy, morbidity and mortality, and pathogenesis. *Annals of Internal Medicine*, **123**, 42–53.

Bruce, I., Gladman, D.D., and Urowitz, M.B. (2000). Premature arterosclerosis in systemkic lupus erythematosus. *Rheumatology Disease Clinics of*

231

North America, **26,** 257-78.

Buchbinder, R. and Hill, C. (2002). Malignancy in patients with inflammatory myopathy. *Current Rheumatology Reports,* **4,** 415-26.

Carey, G. and Gottesman (1978). Reliability and validity in binary ratings. *Archives of General Psychiatry,* **35,** 1454-9.

Cox, D. (1961). Tests of separate families of hypotheses. *Fourth Berkeley Symposium,* **1,** 105-23.

Cox, D.R. and Hinkley, D.V. (1974). Theoretical statistics.

Cox, D. R. (1957). Note on grouping. *JASA,* **52,** 543-7.

Cronbach, L.J., Gleser, G.C., Nanda, H., and Rajaratnam, N. (1972). *The dependability of behavioural measurements.* Wiley: New York.

DeVere, R. and Bradley, M.G. (1975). Polymyositis: its presentation, morbidity and mortality. *Brain,* **98,** 637-66.

Dickey, B.F. and Myer, A.R. (1974). Pulmonary disease in polymyositis/dermatomyositis. *Seminars in Arthritis and Rheumatism,* **14,** 60-76.

Diggle, Peter J., Heagerty, Patrick, Liang, Kung-Yee, and Zeger, Scott L. (2002). *Analysis of longitudinal data.* Oxford University Press.

Farewell, V. T. (1982). A note on regression analysis of ordinal data with variability of classification. *Biometrika,* **69,** 533-8.

Farewell, V. T., Tom, B. D. M., and Royston, P. (2004). The impact of dichotomization on the efficiency of testing for an interaction effect in exponential family models. *JASA.*

Finkelstein, Dianne M. (1986). A proportional hazards model for interval censored failure time data. *Biometrics,* **42,** 845-54.

Fisher, R.A. (1925). *Statistical methods for research workers.* Oliver and Boyd: Edinburgh.

232

Fleiss, J.L. (1971). On the distribution of a linear combination of independent chi squares. *Journal of the American Statistical Association*, **66**.

Fleiss, J.L. (1985). *The design and analysis of clinical experiments*. Wiley: New York.

Fleiss, J.L. and Shrout, P.E. (1978). Approximate interval estimation for a certain intraclass correlation coefficient. *Psychometrika*, **43**, 259–62.

Fleiss, Joseph. L. and Cohen, Jacob. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, **33**, 613–9.

F.W.Miller, Rider, L.G., Chung, Y.-L., Cooper, R., Danko, K., and Farewell, V. (2001). Proposed preliminary core set of measures for disease outcome assessment in adult and juvenile idiopathic inflammatory myopathies. *Rheumatology*, **40**, 1262–73.

Gaylor, D.W. and Hopper, F.N. (1969). Estimating the degrees of freedom for linear combinations of mean squares by Satterthwaite's formula. *Technometrics*, **11**, 691–705.

Gladman, D.D., Urowitz, M.B., and Goldsmith, C.H. (1997). The reliability of the systemic lupus international collaborating clinics/american college of rheumatology damage index in patients with systemic lupus erythematosus. *Arthritis and Rheumatism*, **40**, 809–13.

Gladman, Dafna. D., Farewell, Vernon. T., and Nadeau, Claude. (1995). Clinical indicators of progression in psoriatic arthritis: Multivariate risk model. *Journal of Rheumatology*, **22**, 675–9.

Gladman, D. D., Goldsmith, C. H., Urowitz, M. B., Bacon, P., Bombardier, C., and Isenberg, D (1992). Cross cultural validation of 3 disease activity indices in systemic lupus erythematosus. *Journal of Rheumatology*, **19**,

608–11.

Goggins, William B., Finkelstein, Dianne M., Schoenfeld, David A., and Zaslavsky, Alan M. (1998). A Markov Chain Monte Carlo EM algorithm for analyzing interval-censored data under the Cox proportional hazards model. *Biometrics*, **54**, 1498–507.

Goggins, William B., Finkelstein, Dianne M., and Zaslavsky, Alan M. (1999). Applying the Cox proportional hazards model when the change time of a binary time-varying covariate is interval censored. *Biometrics*, **55**, 445–51.

Hastie, Trevor. and Tibshirani, Robert. (1986). Generaiized additive models. *Statistical Science*, **1**, 297–310.

Haupt, H.M., Moore, W.G., and Hutchins, G.M. (1981). The lung in systemic erythematosus. Analysis of the pathologic changes in 120 patients. *American Journal of Medicine*, **71**, 791–7.

Hay, E.M., Bacon, P.A., Gordon, C., *et al.* (1993). The BILAG index: a reliable and validated instrument for measuring clinical disease activity in systemic lupus erythematosus. *Q J Med*, **86**, 447–58.

Hay, E.M., Black, D., Huddy, A., *et al.* (1994). A prospective study of psychiatric disorder and cognitive function in systemic lupus erythematosus. *Annals of Rheumatic Diseases*, **53**, 298–303.

Horton, Nicholas J., Bebchuk, Judith D., Jones, Cheryl L., *et al.* (1999). Goodness of fit for GEE: an example with mental health service utilization. *Statistics in Medicine*, **18**, 213–22.

Hosmer, D.W. and Lemeshow, S. (1989). *Applied logistic regression.* Wiley.

Joffe, M. M., Love, L.A., and et al, R.L. Leff (1993). Drug therapy of the idiopathic inflammatory myopathies: predictors of response to prednisolone,

azathioprine, methotrexate and comparison of their efficacy. *American Journal of Medicine*, **94**, 379–87.

Johnson, A.E., Gordon, C., Palmer, R.G., *et al.* (1995). The prevalence and incidence of systemic lupus erythematosus (SLE) in Birmingham UK, related to ethnicity and country of birth. *Arthritis and Rheumatism*, **38**, 551–8.

Kalbfleisch and Lawless (1985). The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association*, **80**, 863–71.

Korn, Edward L. and Whitmore, Alice S. (1979). Methods for analyzing panel studies of acute effects of air pollution. *Biometrics*, **35**, 795–802.

Kraemer, Helena Chmura, Periyakoil, Vyjeyanthi S., and Noda, Art (2002). Kappa coefficients in medical research. *Statistics in Medicine*, **21**, 2109–29.

Lagakos, S.W., Sommer, C.J., and Zelen, M. (1978). Semi-Markov models for partially censored data. *Biometrika*, **65**, 311–7.

Liang and Zeger (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13–22.

Liang, K.Y. and Zeger, S.L. (1992). Multivariate regression analyses for categorical data. *Journal of the Royal Statistical Society, Series B.*, **54**, 3–40.

Liang, M.H., Sacher, S.A., Roberts, W.N., *et al.* (1988). Measurement of systemic lupus erythematosus activity in clinical research. *Arthritis and Rheumatism*, **31**, 817–25.

Liang, M. H., Socher, S. A., G., M., and Schur, P. H. (1989). Reliability and validity of six systems for the clinical assessment of disease activity in systemic lupus erythematosus. *Arthritis and Rheumatism*, **32**, 1107–18.

Lin, Lawrence I-Kuei (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, **45**, 255–68.

Mackay, I.R., Taft, L.I., and Cowling, D.C. (1959). Lupoid hepatitis and the hepatic lesions of systemic lupus erythematosus. *Lancet*, **1**, 65–9.

Matthews, David. E. (1984). Efficiency considerations in the analysis of a competing-risk problem. *The Canadian Journal of Statistics*, **12**, 207–10.

McCullagh, P. and Nelder, J.A. (1989). *Generalized linear models*. Chapman and Hall.

Medsger, T.A., Dawson, W.N., and Masi, J.T. (1970). The epidemiology of polymyositis. *American Journal of Medicine*, **48**, 715–23.

Moore, D.S. and Spruill, M.C. (1975). Unified large-sample theory of general chi-squared statistics for tests of fit. *Annals of Statistics*, **3**, 599–616.

Morrow, J., Nelson, L., Watts, R., and Isenberg, D. (1999). *Autoimmune rheumatic disease*. Oxford University Press.

Nelder, J.A. (1954). The interpretation of negative components of variance. *Biometrika*, **41**, 544–8.

Nickerson, Carol A. E. (1997). A note on 'a concordance correlation coefficient to evaluate reproducibility'. *Biometrics*, **53**, 1503–7.

Nived, O., Josen, A., Bengtsson, A., and Sturfeld, G. (2000). High predictive value of SLICC damage index for survival in SLE. *Arthritis and Rheumatism*, **43**, S 250.

Oddis, C.V. and Medsger, T.A. (1989). Current management of polymyositis and dermatomyositis. *Drugs*, **37**, 382–90.

Prentice and Zhao (1991). Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics*, **47**, 825–40.

Rasaratnam, I. and Ryan, P. (1995). Lupus: Advances and remaining challenges. *The Medical Journal of Australia*, **163**, 398–9.

Samanta, A., Feehally, J., Roy, S., Nichol, F. E., Sheldon, P. J., and Walls, J. (1991). High prevalence of systemic disease and mortality in asian subjects with systemic lupus erythematosus. *Annals of the Rheumatic diseases*, **50**, 490–2.

Satten, G.A. (1996). Rank-based inference in the proportional hazards model for interval censored data. *Biometrika*, **83**, 355–70.

Satterthwaite, F.E (1946). An approximate distribution of estimates of variance components. *Biometrics*, **2**, 110–4.

Searle, Shayle. R., Casella, George., and McCulloch, Charles. E. (1992). *Variance components*. John Wiley and Sons, Inc.

Shortall, E., Isenberg, D.A., and Newman, S. (1995). Factors associated with mood and mood disorders in SLE. *Lupus*, **4**, 272–9.

Shoukri, Mohammed. M. (2004). *Measures of interobserver agreement*. Chapman and Hall/CRC.

Shrout, P.E. (1998). Measurement reliability and agreement in psychiatry. *Statistical Methods in Medical Research*, **7**, 301–17.

Shrout, P.E. and Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, **86**, 420–8.

Sigurgeirsson, B., Lindelof, B., Edhag, O., *et al.* (1992). Risk of cancer in patients with dermatomyositis or polymyositis: a population based study. *New England Journal of Medicine*, **326**, 363–7.

Singer, B. (1981). *Sociological methodology*, chapter Estimation of nonstationary Markov chains from panel data. Jossey-Bass.

Singer, B. and Spilerman, S. (1976). Some methodological issues in the analy-

sis of longitudinal surveys. *Annals of Economic and Social Measurement*, **5**, 447–74.

Sitgreaves, R. (1960). Review of intraclass correlation and the analysis of variance by E. A. haggard. *Journal of the American Statistical Association*, **55**, 384–5.

Stoll, T., Stucki, G., Malik, J., *et al.* (1997). Association of the SLICC/ACR damage index with measures of disease activity and health status in patients with systemic lupus erythmatosus. *Journal of Rheumatology*, **24**, 309–13.

Storb, R., Prentice, R.L., and Thomas, E. D. (1977). Marrow transplantation for the treatment of aplastic anaemia: and analysis of the factors associated with graft rejection. *New England Journal of Medicine*, **296**, 61–6.

Stratford, P. (1989). Reliability: consistency or differentiating among subjects? *Physical Therapy*, **69**, 299–300.

Tami, L.F. and Bhasin, S. (1993). Polymorphism of the cardiac manifestations in dermatomyositis. *Clinical Cardiology*, **16**, 260–4.

Tan, E.M., A.S.Cohen, Fries, J.F., *et al.* (1982). The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis and Rheumatism*, **25**, 1271–7.

Taylor, A.J., Wortham, D.C., Burge, J.R., *et al.* (1993). The heart in polymyositis: a prospective evaluation of 26 patients. *Clinical Cardiology*, **16**, 802–8.

Turnbull, B.W. (1976). The empirical distribution function from arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society. Series B*, **38**, 290–5.

Wallance, Daniel. J. (1995). *The lupus book. a guide for patients and their families.* Oxford University Press.

Ware, James H., Lipsitz, Stuart, and Speizer, Frank E. (1988). Issues in the analysis of repeated categorical outcomes. *Statistics in Medicine*, **7**, 95–107.

Watts, R. and Isenberg, D.A. (1989). Pancreatic complications of the autoimmune rheumatic diseases. *Seminars in Arthritis and Rheumatism*, **19**, 158–65.

Wei, G. C. G. and Tanner, M. A. (1990). A monte carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association*, **85**, 699–704.

White (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, **50**, 1–25.

Zeger, Scott L. and Liang, Kung-Yee (1992). An overview of methods for the analysis of longitudinal data. *Statistics in Medicine*, **11**, 1825–39.