ASSESSMENT OF SYSTEMIC LUPUS ERYTHEMATOSUS DIAGNOSES WITHIN QUEBEC'S HEALTH ADMINISTRATIVE DATABASES

Ryan Ng, Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science

Table of contents

Table of figures

Abbreviations

ACR	American College of Rheumatology
ANA	Antinuclear antibody
BMD	Bone mineral density
CI	Confidence interval
DM	Diabetes mellitus
EBV	Epstein-Barr virus
F:M	Female to male
HR	Hazard Ratio
HRT	Hormone replacement therapy
IBD	Inflammatory bowel disease
ICD	International Classification of Diseases
K-M	Kaplan-Meier
LED	lupus érythémateux disséminé
LUMINA	Lupus in Minorities: Nature versus Nature
Med-echo	Maintenance et exploitation des données pour l'étude de la clientèle hospitalière
MS	Multiple sclerosis
NSAID	Non-steroidal anti-inflammatory drug
NHL	Non-Hodgkin's Lymphoma
OA	Osteoarthritis
OR	Odds ratio
P	P – value
PH	Proportional hazards
PPV	Positive predictive value
PROFILE	The Genetic Profile Predicting the Phenotype
RA	Rheumatoid arthritis
RAMQ	Régie de l'assurance maladie du Québec
RR	Risk ratio
SARD	Systemic autoimmune rheumatic disease
SDI	SLICC/ACR damage index
SLE	Systemic lupus erythematosus
SIR	Standardized Incidence Ratio
SLEDAI	SLE Disease Activity Index
SLICC	Systemic Lupus International Collaborating Clinics
SMR	Standardized mortality ratio
UV	Ultra-violet

Abstract

Background: Systemic lupus erythematosus (SLE) is a chronic, relatively uncommon autoimmune disease that has a relapsing-remitting course, with clinical manifestations in various organ systems (cutaneous, renal, and other). To control disease, immunosuppressive drugs are often required. Health administrative databases are useful for studying SLE because of their wide population coverage, and could potentially be used to study SLE incidence, prevalence, clinical manifestations, and medication use. However, because the diagnoses in these administrative databases are not necessarily clinically confirmed, SLE case ascertainment is a methodological challenge. First, some of the methodological issues were examined in this thesis. Second, clinical manifestations and the association between early antimalarial drug use and future renal manifestations were examined in a cohort of SLE patients.

Methods: The initial SLE case definition was a previously-used algorithm that identified subjects as having SLE if they met one of the following criteria: one SLE hospital discharge code, one rheumatologist SLE claim and/or two SLE non-rheumatologist claims at least eight weeks apart but within two years. Alternative algorithms were formed by modifying one or more of the initial algorithm's parameters. Incidence and prevalence estimates were determined using each alternative algorithm and compared to the initial estimates. The effect of using different data period lengths for detecting patients was also examined.

Kaplan-Meier (K-M) analyses were performed to assess documentation of clinical SLE manifestations and use of selected immunosuppressant medications, within an incident SLE cohort identified by the initial algorithm (described above). The

observation interval began four years prior to SLE diagnosis and continued up to eight years after SLE diagnosis. Cox proportional hazards regression analyses were used to examine the association between early antimalarial drug use and renal manifestations.

Results: With the initial algorithm, the 1998 yearly incidence was 6.0 cases per 100,000 (95% confidence interval (CI), 5.5–6.6). When parameters from the initial algorithm were changed, the 1998 incidence varied to between 4.4 and 7.4/100,000. The prevalence also changed from 65.5/100,000 (95% CI: 63.7–67.4) with the initial algorithm, to between 47.8–79.1/100,000 with the alternate algorithms. When the length of the data period changed from fifteen years to five years, the 2001 yearly incidence was overestimated by 38.3% (5.7/100,000 initially and 7.9/100,000 with only five years of data) and the prevalence was underestimated by 29.9% (the new estimate being 46.0/100,000, 95% CI: 44.4–47.5).

Over-all, 66.2% (95%CI: 63.4–68.9%) of incident patients (within the SLE cohort assembled using the initial algorithm) had evidence of at least one SLE manifestation within the period under examination. The most common manifestation was cutaneous involvement, present in 30.0%. Within the sub-cohort of incident SLE patients covered by RAMQ drug insurance, 87.2% (95% CI: 84.2–90.3%) had received at least one of the medications under study, by the end of the study interval. No association was found between early antimalarial drug use and subsequent renal manifestations.

Conclusion: Varying the case definition and data period can change incidence and prevalence estimates considerably, so all features, including the time period in which the data spans, should be selected carefully and explicitly stated. The majority of incident SLE patients had evidence of SLE manifestations or used medications which

would provide possible confirmation of SLE case status. This additional information can be used in future health services administrative database research to understand SLE, and help compensate for the databases' lack of clinical confirming data.

Résumé

Contexte : Le lupus érythémateux disséminé (LED) est une maladie auto-immune chronique relativement peu commune. L'évolution de cette maladie est décrite en phases de poussées et de rémissions et ses manifestations cliniques touchent plusieurs organes, dont la peau, les reins, etc. L'utilisation de médicaments immunosuppresseurs est souvent nécessaire pour contrôler le LED. Les banques de données administratives du domaine de la santé s'avèrent utiles pour étudier le LED, car elles couvrent une vaste population et elles pourraient être utilisées pour étudier l'incidence, la prévalence, les manifestations cliniques et l'utilisation de médicaments chez des patients atteints du LED. Toutefois, comme les diagnostics présents dans ces bases de données administratives n'ont pas nécessairement de confirmation clinique, la détermination des cas de LED représente un défi d'ordre méthodologique et certains de ces problèmes méthodologiques font l'objet de la présente thèse.

Méthodologie : L'algorithme initial de définition de cas de LED a déjà été utilisé pour identifier des sujets atteints de LED s'ils répondaient aux critères suivants : un code de congé d'hôpital de LED, une réclamation d'un rhumatologue pour le LED et/ou deux réclamations par un médecin autre qu'un rhumatologue pour le LED séparées d'au moins huit semaines, mais dans un intervalle de deux ans. D'autres algorithmes ont été créés en modifiant un paramètre ou plus de l'algorithme initial. Des estimations d'incidence et de prévalence ont été obtenues grâce à chaque algorithme créé et ces valeurs ont été comparées aux estimations initiales. L'effet de l'utilisation de périodes de données de différentes longueurs sur la détection des patients a également été examiné.

Des analyses Kaplan-Meier (K-M) ont été faites pour évaluer la documentation des manifestations cliniques du LED et l'utilisation de médicaments immunosuppresseurs spécifiques au sein d'une cohorte incidente de patients atteints de LED identifiés par l'algorithme initial (décrit plus haut). L'intervalle d'observation a débuté quatre ans avant le diagnostic de LED et s'est poursuivi jusqu'à huit ans après le diagnostic. Des analyses utilisant le modèle de régression à risques proportionnels de Cox ont servi à examiner l'association entre l'utilisation précoce d'antipaludiques et les manifestations rénales.

Résultats : Avec l'algorithme initial, l'incidence annuelle de LED en 1998 était de 6,0 cas pour 100 000 habitants (95 % d'intervalle de confiance (CI), 5,5-6,6). En changeant les paramètres de l'algorithme initial, l'incidence en 1998 a varié entre 4,4 et 7,4 pour 100 000. La prévalence a passé de 65,5 pour 100 000 (95 % CI : 63,7–67,4) avec l'algorithme initial à entre 47,8–79,1 pour 100 000 avec les autres algorithmes. En modifiant la longueur des périodes de données de quinze à cinq ans, l'incidence annuelle en 2001 était surestimée par 38,3 % (5,7 pour 100 000 initialement et 7,9 pour 100 000 avec seulement cinq ans de données) et la prévalence était sous-estimée par 29,9 % (le nouvel estimé étant 46,0 pour 100 000, 95 % CI : 44,4–47,5).

Dans l'ensemble, 66,2 % (95 % CI : 63,4–68,9 %) des patients incidents au sein de la cohorte de patients atteints de LED assemblée grâce à l'algorithme initial montraient au moins une manifestation de LED au cours de la période évaluée. La manifestation clinique la plus commune était l'atteinte cutanée, présente à 30,0 %. Au sein d'une sous-cohorte de patients incidents atteints de LED couverts par la RAMQ, 87,2 % (95 % CI : 84,2–90,3 %) ont reçu au moins un médicament à l'étude avant la fin

de l'intervalle étudié. Aucune association n'a été trouvée entre l'utilisation précoce d'antipaludiques et les manifestations rénales subséquentes.

Conclusion : La variation de la définition de cas et de la période de données peut modifier considérablement les estimations d'incidence et de prévalence. Ainsi, tous les paramètres, y compris la période de temps pour laquelle les données sont recueillies, devraient être choisis avec précaution et énoncés clairement. La majorité des patients incidents atteints de LED montrent des manifestations de LED ou consomment des médicaments qui pourraient offrir une confirmation potentielle des cas de LED. Ces informations supplémentaires pourront être utilisées pour des études futures sur les bases de données des services de soins de santé afin de mieux comprendre le LED et de permettre de compenser pour le manque de confirmation des données cliniques des bases de données.

Acknowledgements

I would like to thank my supervisor Dr. Elham Rahme and my co-supervisor Dr. Sasha Bernatsky for all the guidance and support they provided throughout my MSc studies. Their instruction has enriched my MSc experience and increased my desire to pursue a career in epidemiology. I would also like to thank Hacène Nedjar for all his help in teaching me SAS programming and helping troubleshoot my code. Finally, my research would not have been possible without the financial contributions provided by the Canadian Institutes of Health Research (CIHR), the Canadian Arthritis Network (CAN) and Mitacs (in partnership with Pfizer).

Contributions of Authors

The two manuscripts presented within are works-in-progress for in preparation for submission. Ryan Ng, the author of this thesis, was responsible for designing the studies, managing the data, analyzing the results and writing up the manuscripts. Dr. Elham Rahme and Dr. Sasha Bernatsky were both involved with study design and manuscript revisions.

Chapter 1: Introduction

Introduction

Systemic lupus erythematosus (SLE) is a relatively uncommon, systemic autoimmune disease with a relapsing-remitting nature that can manifest in many organs such as the skin, joints, and kidneys, often simultaneously. SLE's relatively low prevalence, coupled with its clinical complexity and varying manifestations, complicates its study. One resource available to study SLE outcomes in Quebec are health administrative databases, like the Maintenance et exploitation des données pour l'étude de la clientèle hospitalière (Med-echo) hospitalization discharge database, the Régie de l'assurance maladie du Québec (RAMQ) physician claims database, and the RAMQ prescription drug database. Each database contains reimbursement data (such as dated physician visits and hospitalizations) for all Quebec permanent residents, so when these data are linked with demographic data, they can provide information that may be used to estimate the incidence and prevalence of chronic diseases by age groups, sex and geographic regions.

However, methodological challenges do exist with these databases, such as how to ascertain cases of disease. These challenges exist for every disease because these databases were designed for administrative purposes, so they do not have clinical data that can verify cases on their own. Many different algorithms have been used for different diseases, and these algorithms are currently in development by many groups including the Public Health Agency of Canada[1]. Some algorithms have been validated for other diseases, but relatively little work has been done to evaluate SLE case ascertainment in Quebec.

An additional challenge for SLE case ascertainment is its relapsing-remitting nature, meaning patients will have periods of high disease activity followed by periods where disease activity is minimal or absent. When SLE is in remission, patients are presumably less likely to receive medical care, so they are less likely to be captured by health administrative databases. Hence for diseases like SLE, the period of time used to identify SLE cases might affect the number of cases ascertained. Theoretically, the more years of data, the greater the chance of ascertainment, but because Canadian administrative data often dates back only to the 1980's, accessing many years of data can be much more costly than accessing only more recent data. It is thus of interest to determine the length of time needed to optimally ascertain cases of SLE.

Health administrative data are continually collected for ongoing physician and hospital reimbursement, which also makes these databases attractive for longitudinal study. These databases have been used to study SLE health care use patterns and associated co-morbidities, but they have not been used to describe SLE manifestations or drug use patterns over time. Examining SLE manifestations and medication use in this manner could help clinicians and researchers better understand the prognosis of SLE, and perhaps guide long-term therapeutic strategies. The association between early SLE treatment and SLE manifestations can also potentially be analyzed with regression analyses.

This thesis attempts to examine some of the methodological challenges of using administrative data for rheumatic disease surveillance and research.

Objectives

This thesis project had four objectives. From a methodological standpoint, there were two objectives. These were to compare incidence and prevalence estimates based on (i) different case ascertainment algorithms, and (ii) various lengths of time for case ascertainment. From an analytical standpoint, there were also two additional objectives: (iii) to examine longitudinally the development of SLE manifestations and the use of SLE medications in an incident SLE cohort, and (iv) to assess the potential association between early SLE treatment and renal manifestations.

Chapter 2: Literature review

This literature review will present an overview of published epidemiological SLE research and the role of health administrative databases in research, divided in two parts. The first part discusses the epidemiology of SLE, focusing on: incidence, prevalence, risk factors, manifestations, treatments, morbidity and mortality. The second part discusses health administrative databases beginning with an overview; followed by sections on their limitations and strategies to handle these limitations; and ending on SLE-specific health administrative research.

SLE epidemiology

SLE incidence and prevalence

Many studies have estimated SLE incidence and prevalence, and 43 of these studies were identified by three recent reviews, Danchenko *et al.* (2006), Lim and Drenkard (2008) and Pons-Estel *et al.* (2010)[2-4], two of which are Canadian (ten studies are shown in Table 2.1; see Appendix A for the full list). The first study by Peschken *et al.* (2000) identified Manitoban SLE cases from a regional arthritis center database and the medical records of all Manitoban rheumatologists, hematologists, nephrologists and general internists who had an SLE patient between 1980 and 1996[5]. The other study by Bernatsky *et al*. (2007) measured Quebec SLE incidence and prevalence from its hospital discharge database and physician claims databases via case ascertainment algorithms using Bayesian latent class models suitable for non-gold standard diagnostic tests[6].

Author(s)	Sources for SLE case ascertainment	Location	Yearly Incidence and 95% CI (/100,000)		Prevalence and 95% CI (/100,000)	
Siegel *et al*. (1970)	Hospital Files	Jefferson County, USA	9.9	--	5.73	--
Amor *et al*. (1983)	Physician survey	France	0.1	--	--	--
Nossent (1992)	Patient records	Curacao	4.6	[0.4, 8.8]	47.0	[34.1, 51.1]
Iseki *et al*. (1994)	Tokutei-Sikkan program applicants, medical association referral, author referral, and dialysis	Okinawa, Japan	0.9	--	37.7	--
Hopkinson *et al*. (1994)	Hospital physician survey, immunology registry and lab, renal database, inpatient medical records, and acute psychiatric admissions	Nottingham, UK	4.0*	[2.4, 5.7]	24.7*	[20.7, 28.8]
Johnson *et al*. (1995)	Physician referral, lupus patient groups, rheumatology and hospital discharge database, and immunology laboratories	Birmingham and Slihull districts, UK	3.8*	[2.5, 5.1]	27.7*	[24.2, 31.2]
Voss *et al*. (1998)	Inpatient and outpatient registry, GP and specialist referral and university autoimmune test database	Funen county, Denmark	2.5	[1.8, 3.3]	21.7	[17.3, 26.8]
Peschken *et al*. (2000)	Specialist and family physician personal patient databases	Manitoba, Canada	--	--	22.1	[13.2, 32.4]
Chakravarty *et al*. (2000)	Hospital inpatient database	Pennsylvania, USA	--	--	149.5	[146.9, 152.2]
Nightingale *et al*. (2006)	General Practice Research Database, medical records and prescription records	United Kingdom	3.0	[2.7, 3.3]	40.7	[37.6, 43.8]
Bernatsky *et al*. (2007)	Physician claims and hospital discharge abstract administrative databases	Quebec, Canada	3.0	[2.6, 3.4]	44.7	[37.4, 54.7]

Table 2.1: *Previous SLE incidence and prevalence studies.*

Twenty-eight studies reported yearly SLE incidence estimates that ranged from 0.1–18.5 cases per 100,000 individuals[7, 8] with the majority (78%) of studies having a yearly incidence between 1 and 9/100,000 (Appendix B). Thirty-eight studies determined prevalence with estimates ranging from 5.7–149.5/100,000[7, 9] with 69% of the examined studies reporting a prevalence estimate between 20 and 70/100,000 (Appendix C). There are many factors that explain these variations, and they include: sex, age, race/ethnicity, geography, study design and risk factors.

Sex

The yearly incidence and prevalence of SLE were much higher in females than males. The yearly incidence estimates for females ranged from 2.5–14.1/100,000 compared to 0.4–2.2/100,000 in males[10-12]. Prevalence estimates for females ranged from 12.5–131.5/100,000 compared to 3.4–24.8/100,000 in males[6, 13-15]. As shown by

these estimates, SLE is a predominantly female disease, a feature common to many other autoimmune rheumatic diseases[16].

Age

Twenty-two studies (48%) examined incidence and prevalence by age groups and each found considerable variation; however, direct comparisons between studies are difficult because each study used different age categories. The observed trend for women was an increase in incidence from birth until the early- to mid-twenties where there was plateauing[17-20]. This plateau lasted until the mid- to late-forties, where incidence started decreasing again. In eight of the studies, the average age of diagnosis was reported, with estimates ranging from 31.4–46.3 years[11, 21]; however, in two other studies, the average age at diagnosis was much higher. In a study by Bernatsky *et al*. in 2007, SLE female incidence was highest in the 45-64 age group[5]; in another study by Somers *et al.* in 2007, SLE incidence was highest in the 50-54 age group[21]. Some of these differences across studies may represent differences in study design; for example, Somers *et al.* ascertained cases from medical records, which might represent a biased sampling reference with respect to older age (compared to the baseline population).

Detecting age trends in males has been more difficult because relatively very few male cases are affected by SLE. Many studies did not report male incidence by age group[22, 23], and for those that did with the exception of two, trends were not clearly apparent[11, 13, 24]. In the Bernatsky *et al.* (2007) and Somers *et al.* (2007) studies which had a large number of SLE cases (3825 and 1638, respectively), male yearly incidence was the highest in older age groups (65+ years and 70-74 years, respectively)[6, 21]. As mentioned earlier, some differences may be due to study design.

Comparing the age at which SLE was diagnosed between sexes, amongst the 43 studies examined, males were diagnosed between 3.6 and 23.4 years later than women were[11, 19, 21, 23, 25]. This age difference may represent a true biologic phenomenon, although other hypotheses are possible, such as: males seek medical care less often than females, or that physicians may be less likely to consider the possibility of SLE in males given that SLE is such a predominantly female disease.

Female to male ratios are often used to summarize age group trends. Amongst pediatric SLE patients (less than 19 years of age), the female to male ratio is about 2:1[26]. During child-bearing years (ages 15–44), the sex ratio jumps to about 9:1[27], with some studies finding a sex ratio as high as 14.4:1[28]. After menopause, within middle-aged onset SLE patients (after the age of 44), the female to male ratio decreases back to 2:1[29].

Race/ethnicity

SLE incidence and prevalence also varies across race/ethnicity groups (Appendix B and C). American studies found SLE incidence to be 2.6–3.2 times higher in African-Americans compared to Caucasians[7, 22, 30]. While this increased risk can be attributable to race/ethnicity differences, inequalities in health care accessibility and affordability have also been suggested[4]. Still, ethnicity differences were found in a UK-based study, a country with a public health care system designed to reduce health care accessibility and affordability inequalities[23]. In this study, SLE prevalences in Afro-Caribbean and Asian populations were 10.2 and 2.4 times higher than in the Caucasian population. These results were duplicated one year later in another UK-based study by Johnson *et al.* (1995), which found yearly incidence in Afro-Caribbean and Asian populations to be 5.1 and 6.5 times higher compared to the local Caucasian population[31]. Peschken *et al.*

(200) have also found yearly incidence to be higher among Manitoban aboriginals versus the rest of the Manitoban population[5].

Geography

SLE incidence and prevalence have also been shown to vary by geographic location, which includes a prevalence gradient with, rather surprisingly, a lower incidence in Africa versus North America and Europe[32]. Some variation has been attributed to different population composition (sex, age and genetic), but the lack of health care infrastructure, in particular, the lack of diagnostic tests to detect SLE has also been suggested as a factor (accessibility, affordability, types of health care providers and SLE disease education are other infrastructure factors)[3]. Another hypothesis is individuals in Africa develop greater resistance to autoimmune diseases because of an increased exposure to malaria and other parasitic infections[32]. A related theory is the higher use of antimalarials to treat malaria may simultaneously be treating early-stage SLE in those with both conditions[33].

Study design

Aspects of study design, such as the population source, the SLE case definition and time-related issues, may also cause variations[3, 34]. The source population is important because it determines the type of SLE cases selected. Case ascertainment from hospital sources like hospital records, hospital discharge databases and hospital registries capture serious, well-defined SLE cases; however, reliance solely on these sources will miss milder, non-hospitalized cases. One study of undiagnosed SLE cases (defined as: individuals diagnosed with SLE after testing positive to a mailed, screening questionnaire and subsequent diagnostic testing) found the overall female SLE

prevalence in a British community to be four times higher than the diagnosed female SLE prevalence, which was based on medical records (200/100,000, 95% CI: 80–412 versus 54/100,000, 95% CI: 47–62)[35]. Because of the complexity of SLE symptoms, case ascertainment by self-report is problematic. Physician surveys can detect outpatient cases, but are subject to recall bias[3]. Health administrative databases cover the entire population, but they rely on patient-initiated contact with the health care system[36]. Of the 43 studies examined by Danchenko *et al.* (2006), Lim and Drenkard (2008) and Pons-Estel *et al.* (2010)[2-4], 53% of them used at least two data sources to select cases, which allows missing case to be estimated through capture-recapture methods[37].

Because SLE cannot be diagnosed with a single definitive diagnostic test, deciding on a case definition is difficult. Past studies have defined SLE cases by diagnostic criteria, self-report, case ascertainment algorithms, physician diagnosis or other validated methods[6, 14, 31, 38]. While each definition type has advantages, different definitions cause comparability issues. Many studies now use the American College of Rheumatology (ACR) criteria to define SLE (63% of the 43 studies in Appendix A used the ACR criteria). The ACR criteria were first devised in 1971, but as clinical understanding of SLE has improved, it has undergone two revisions[39]. Presently, the 1997-revised ACR criteria is used, and it identifies a person as having SLE if he or she exhibits at least four of eleven criteria (Table 2.2), either concurrently or consecutively. The ACR criteria for SLE case definition were developed for randomized control trials and do not constitute a gold standard, so it may lack sensitivity in some studies[40]. For example, the ACR criteria may under-detect patients with early stage SLE, lupus nephritis and neuropsychiatric manifestations[41]. The over-all specificity and sensitivity of these criteria are 95% and 85% respectively[42], and milder cases are also more likely to be missed. Alternative

criteria like the Cleveland Clinic weighted criteria, the Boston weighted criteria, the St. Thomas alternative criteria and classification trees[41, 43] exist, but the ACR criteria remains by far the most often-used. Even today, the ACR criteria are still changing; currently, the Systemic Lupus International Collaborating Clinics (SLICC) is revising the 1997 ACR criteria[44].

Criterion	Criterion Definition
1. Malar Rash	Fixed erythema, flat or raised, over the malar eminences, tending to spare the nasolabial folds
2. Discoid rash	Erythematosus raised patches with adherent keratotic scaling and follicular plugging; atrophic scarring may occur in older lesion
3. Photosensitivity	Skin rash as a result of unusual reaction to sunlight, by patient history or physician observation
4. Oral Ulcers	Oral or nasopharyngeal ulceration, usually painless, observed by physician
5. Nonerosive arthritis	Involving two or more peripheral joints, characterized by tenderness, swelling, or effusion
6. Pleuritis or pericarditis	Pleuritis - convincing history of pleuritic pain or rubbing heard by a physician or evidence of pleural effusion, OR Pericarditis - documented by electrocardiogram or rub or evidence of pericardial effusion
7. Renal disease	Persistent proteinuria > 0.5 grams per day or greater than +++ if quantification not Cellular casts - may be red cell, hemoglobin, granular, tubular or mixed
8. Neurologic disorder	Seizures - in the absence of offending drugs or known metabolic derangements; e.g. uremia, ketoacidosis, or electrolyte imbalance, OR Psychosis - in the absence of offending drugs or known metabolic derangements; e.g. uremia, ketoacidosis, or electrolyte imbalance
9. Hematologic disorder	Hemolytic anemia: with reticulocytosis, OR Leukopenia: < 4,000/mm^3 on at least 2 occasions, OR Lymphopenia: < 1,500/mm^3 on at least 2 occasions, OR Thrombocytopenia: < 100,000/mm^3 in the absence of offending drugs
10. Immunologic disorder	Anti-DNA - antibody to native DNA in abnormal titer, OR Ant-Sm - presence of antibody to Sm nuclear antigen, OR Positive finding of antiphospholipid antibodies on: i) An abnormal serum level of IgG or IgM anticardiolipid antibodies ii) A positive test result for lupus anticoagulant using a standard method, or, iii) A false-positive test result for at least 6 months confirmed by Treponema pallidum immobilization or fluorescent treponemal antibody absorption test
11. Positive Antinuclear Antibody	An abnormal titer of antinuclear antibody by immunofluorescence or an equivalent assay at any point in time and in the absence of drugs

Table 2.2: *1997 Update of the 1982 American College of Rheumatology Revised Criteria for Classification of Systemic Lupus Erythematosus.* The proportions displayed are from Tan *et al.* (1982)[39].

There are also time-related methodology issues when detecting cases. Period effects can affect incidence and prevalence estimates. Over time, improvements in SLE diagnosis from earlier physician recognition and improved diagnostic tests such as the antinuclear antibody (ANA) test[39] increased SLE case detection, resulting in trends like the three- to seven-fold increase seen in American SLE incidence across 1950–1992[45]. The duration of time used to detect SLE cases is also important. If too short a time is used, some SLE cases might be missed because of SLE's relapsing-remitting nature.

Risk factors

Finally, across study populations, the level and distribution of SLE risk factors (discussed next) may vary, affecting incidence and prevalence estimation.

SLE risk factors

Sex

As indicated by the large proportion of women with SLE compared to men, sex is one risk factor, and a few biological hypotheses have attempted to explain this phenomenon, such as: the *sex hormone hypothesis*, the *sex chromosome hypothesis* and the *intrauterine selection hypothesis*[46]. The *sex hormone hypothesis* is based on the differences in sex hormone levels (estrogen, testosterone and prolactin) seen between sexes, which increases SLE risk in women. Laboratory studies have shown that estrogen induces IgG and IgM autoantibody formation[47] and prolactin improves the survival of autoreactive B cells[48], two effects that increase the sensitivity of the autoimmune system. Conversely, testosterone may be protective through anti-dsDNA antibody production suppression[49]. Epidemiological studies have also examined risk of SLE related to menarche, menopause, oral contraceptive use and hormone replacement

therapy (HRT) use. One study found girls who reached menarche early (at age ten or younger) were at double the risk of SLE (risk ratio (RR) = 2.1, 95% CI: 1.4–3.2) versus those who reached menarche at age twelve[50]. The study also found that any oral contraceptive use increased SLE risk by 50% (RR = 1.5, 95% CI: 1.1–2.1), and postmenopausal HRT use was associated with a 90% increased risk (RR = 1.9, 95% CI: 1.2–3.1). This hypothesis is also supported by a randomized trial which showed taking HRT after menopause increased the risk of experiencing mild to moderate SLE flares by 34% (95% CI: 1.07, 1.66)[51]. While this hypothesis explains why more women may develop SLE, it does not explain why SLE activity can be more severe in men[46]. The *sex chromosome hypothesis* attempts to explain this observation by suggesting that differences in expression of immunologically-related genes on the X chromosome like the C40 ligand and interferon-related genes affect SLE expression and its severity[16].

A third hypothesis is the *intrauterine selection hypothesis*, which suggests a negative selection of male fetuses at risk for SLE occurs during pregnancy. This hypothesis is based on evidence showing that a lower proportion of males are born to mothers with SLE compared to non-SLE mothers[52]. Two biologically-possible explanations are: the Y chromosome contains a lethal gene which affects only male fetuses, or male fetuses at risk for SLE are more likely to be perceived as antigenic and attacked by the mother's immune system[46].

Genetics

There is evidence of genetics as a risk factor. Twin studies have shown a 25–69% SLE concordance among monozygous twins versus a 2% SLE concordance between dizygous twins[53, 54]. A study looking at SLE familial aggregation found up to 10% of SLE

patients also have a family member with SLE[55]. Linkage studies have identified genes that may increase SLE risk, such as Fcγ receptor genes , the programmed cell death 1 gene and the human leukocyte antigen gene[56].

Environmental risk factors

The Epstein-Barr virus (EBV), smoking, silica and ultra-violet (UV) radiation are environmental risk factors. EBV has been linked to SLE in serologic response studies with one study showing an increased risk for SLE when an EBV capsid antigen serological response occurs (odds ratio (OR) = 49.9, 95% CI: 9.3–1025)[57]. Smoking as a risk has been thoroughly investigated, and a meta-analysis by Costenbader *et al.* (2004) found current smokers had higher odds of developing SLE than non-smokers (OR = 1.50, 95% CI: 1.09–2.08), but former smokers compared to nonsmokers did not have higher odds of developing SLE (OR = 0.98, 95% CI: 0.75–1.27)[58]. Silica has been examined as a risk factor because of its stimulatory effects on the immune system. One case-control study of occupational silica exposure found a significant association between SLE and medium (OR = 2.1, 95% CI: 1.1–4.0) and high silica exposure (OR = 4.6, 95% CI: 1.4–15.4) compared to no silica exposure (in this study, medium exposure was defined as exposure to grinding glass, plastic or other materials; or spraying/sanding enamel; high exposure was exposure to sandblasting, mining, masonry, and pottery/ceramic manufacturing) [59]. Certain UV radiation has been shown to exacerbate and induce skin lesions In SLE patients (UV-A: 320–340 nm), but other UV wavelengths (UV-B: 340–400 nm) are therapeutically beneficial[60].

Other potential SLE risk factors and protective factors

Other potential SLE risk factors include solvents, pesticides, heavy metals, alcohol consumption and breastfeeding. There has only been limited research on solvent exposure[61], but one recent study showed an association between occupational exposures to solvents like nail polish and SLE development (OR = 10.2, 95% CI: 1.3–81.5)[62]. Another study examining self-reported exposure to solvents, pesticides and mercury found an association between SLE and mixing pesticides during agricultural work (OR = 7.4, 95% CI: 1.4–40) and with mercury exposure (OR = 3.6, 95% CI: 1.3–10.0)[63]. Alcohol consumption is a controversial protective factor. One meta-analysis found a protective effect of moderate alcohol drinking in SLE patients receiving treatment for up to ten years (OR = 0.72, 95% CI: 0.55–0.95), but this association was not seen in patients only receiving treatment for up to five years[64]. Breastfeeding may also be protective; a 2002 Carolina study found breastfeeding had a protective effect against SLE development (OR = 0.6, 95% CI: 0.4–0.9)[65]. However, a 2007 study found a null effect of breastfeeding, even as the number of weeks of breastfeeding increased[50].

When SLE and its risk factors are better understood, SLE diagnosis is likely to improve. The next section discusses the wide array of SLE symptoms and manifestations that can be expressed in a patient, and the difficulty of diagnosing SLE.

SLE symptoms and manifestations

While SLE is known for systemic inflammation, a malar (or 'butterfly' rash) and renal lesions, all caused by auto-antibodies, patients can also have a wide constellation of other manifestations[4]. SLE may present itself ambiguously as joint pain, inflammation and fatigue, but it can also affect organ systems like the lungs, central nervous system and especially the kidney (Table 2.3). The number and severity of these symptoms can

vary over time because of SLE's relapsing-remitting nature. The length of remission has been reported to be about 2.3 years (range of 1.1–5.7 years)[66]. While SLE diagnosis has improved, because of its complexity and symptoms common to other diseases, it may take two years to confirm SLE diagnosis after symptom onset[67]. The presentation of SLE symptoms and manifestations has been shown to vary by patient characteristics, such as: sex, race/ethnicity, age of onset and progression of the disease.

Organ system	Proportion of patients (%)	Symptoms
Constitutional	50 – 100	fatigue; fever (in the absence of infection); weight loss
Musculoskeletal	73	arthritis; arthralgia; myositis
Cutaneous	62 – 67	butterfly rash; photosensitivity; mucous membrane lesion; alopecia; Raynaud's phenomenon; purpura; urticaria; vasculitis
Renal	16 – 38	Hematuria; proteinuria; casts; nephritic syndrome
Gastrointestinal	36	Nausea; vomiting; abdominal pain
Pulmonary	7 – 23	Pleurisy; pulmonary parenchyma; pulmonary hypertension
Cardiac	12 – 21	Pericarditis; endocarditis; myocarditis
Reticuloendothelial	18	Lymphadenopathy; splenomegaly; hepatomegaly
Hematologic	15	Anemia; thrombocytopenia; leukopenia
Neuropsychiatric	2-21	Psychosis; seizures; organic brain syndrome; transverse myelitis; cranial neuropathies; peripheral neuropathies

Table 2.3: *SLE manifestations (adapted from 'Guidelines for referral and management of systemic lupus erythematosus' from the American College of Rheumatology's ad hoc committee on systemic lupus erythematosus guidelines (1999)[42]).*

Sex

Studies have observed differences in the types of SLE manifestations expressed between sexes. A review by Lu *et al.* (2010) found renal manifestations such as nephritis and proteinuria consistently occurred more frequently in males[46]. They also found discoid lesions and acute lesions were more common in males, but not malar rash. Hematological manifestations like thrombocytopenia, hemolytic anemia and lymphopenia were also more common in males, as well as serositis and neurological involvement. The ACR criteria less common in men versus women were photosensitivity

and mucosal ulcers. From the studies examined, there was no clear consensus about whether arthritis was more or less common in males versus females. For immunologic criteria, anti-Ro and anti-LA were less prevalent in men, but anti-dsDNA, anti-Sm, anticardiolipin and lupus anticoagulant antibodies were more prevalent.

Race/ethnicity

The types of manifestations expressed also vary by race/ethnicity as seen in the PROFILE (The Genetic Profile Predicting the Phenotype) cohort, a multiethnic cohort of Caucasians, African-Americans and Hispanics who were diagnosed with SLE at most ten years before enrollment[68]. Within this cohort, the ACR manifestations were compared across race/ethnicity. All three groups had somewhat similar proportions of arthritis, neurologic disorders, and positive ANA, but differed considerably on the eight other criteria. Compared to Hispanics and Caucasians, African-American SLE patients were less likely to have malar rash, photosensitivity, or oral ulcers, but more likely to have discoid rash. Caucasians were less likely to have renal manifestations, serositis, cytopenia, and immunologic than the other groups. ACR manifestations in Hispanics were always similar or intermediary to one of the other two groups.

In the American SLE LUMINA (Lupus in Minorities: Nature versus Nature) cohort, the initial presenting ACR manifestations were compared between the same three racial/ethnicity groups (Caucasians, African-Americans and Texan Hispanics)[69]. In this study, the only statistically-similar initial manifestation was neuropsychiatric disorder. Arthritis and positive ANA presented less commonly in Caucasians compared to Texan Hispanics and African Americans. As seen in the PROFILE cohort, compared to Texan Hispanics and Caucasians, African Americans were less likely to present with malar rash

photosensitivity and mucosal ulcers, but more likely to present with discoid. Similar results between studies were also seen in Caucasians where they were less likely to present with serositis, renal manifestations, hematologic manifestations and immunologic manifestations compared to Texan Hispanics and African Americans.

In the 1000 Canadian Faces of Lupus cohort, there was variation in ACR manifestations at the baseline visit between Caucasians, Aboriginals, Afro-Caribbeans and Asians[70]. Across groups, positive ANA and neurologic disorders were similar. Compared to Caucasians, Asians were less likely to have photosensitivity, arthritis and serositis, but more likely to have renal, hematologic and immunologic manifestations. Aboriginals, compared to Caucasians, were less likely to have malar rash, oral ulcers, photosensitivity and hematologic manifestations.

Age

Pediatric SLE cases (<19 years of age), which make up 15–20% of total SLE cases[71], and late-onset patients (consisting of middle-onset (ages 45–64) and elderly onset (ages 65+)), which consist of 12–18% of total cases[72], both have different presenting manifestations in comparison to young-adult onset SLE patients (ages 19–44). In the Euro-Lupus cohort, pediatric SLE cases were more likely to present with malar rash and nephropathy compared to adults[73]. In another study, compared to an adult population, pediatric patients were more likely to present with nephropathy, fever and lymphadenopathy and later develop malar rash and chorea[74]. Another study found neuropsychiatric disorder was the most common pediatric symptom over glomerulonephritis (95% vs. 55%)[75]. Other more prevalent manifestations in pediatric

cases include thrombocytopenia and hemolytic anemia, and overall, pediatric cases presented with more serious organ involvement[67].

The manifestations of late-onset SLE are also different. At presentation, the Euro-Cohort study found malar rash, arthritis and nephropathy were less likely to present in this group compared to younger SLE patients, and Sicca syndrome was more common as the disease progressed[73]. Late-onset patients were also more likely to present with benign, constitutional symptoms like arthralgia, weakness, fatigue, myalgia, weight loss, pyrexia and loss of cognitive function[72], resembling patients with drug-induced SLE (an SLE-mimicking autoimmune response due to chronic use of certain drugs like hydralazine), primary Sjogren's syndrome or polymyalgia rheumatica[67].

The age of diagnosis also differs by racial/ethnicity. Caucasians are more likely to have late-onset SLE[72]. In the 1000 Faces cohort, the average SLE-onset age was 33 ± 14 years (standard deviation) in Caucasians, compared to 34 ± 12.5 years in Aboriginals, 30 ± 10.6 years in Afro-Caribbeans and 25 ± 11.7 years in Asians[70]. In three other studies, the age of onset in Caucasian females was 4.6–7.0 years later than Afro-Caribbean females (41.7 vs. 35.5 years (P-value (P) = 0.005)[76]; 39.8 vs. 35.2 years ($P < 0.05$) [30]; 33.0 vs. 26.0 years ($P < 0.05$))[31]. Two studies of Aboriginal populations in Canada and Australia found the average age of onset to be 31 and 28.2 years, respectively[5, 77].

SLE treatments

There are many drug therapies available for SLE treatment, including: non-steroidal anti-inflammatory drugs (NSAIDs), antimalarials (hydroxychloroquine and chloroquine), corticosteroids (prednisone and methylprednisolone),

immunosuppressants (classically cyclophosphamide, methotrexate, mycophenolate mofetil and azathioprine) and biologics (rituximab, belumimab) (Table 2.4). While beneficial, these treatments are not without adverse effects; examples of adverse effects include: corticosteroids with osteoporosis, avascular necrosis and infection[42]; hydroxychloroquine with macular damage [42]; and cyclophosphamide with teratogenicity, bone marrow suppression and infection[78].

Symptom	Treatment
Cutaneous	Sunscreen, topical steroids, hydroxychloroquine, prednisone, azathioprine and mycophenolate mofetil (for refractory cases)
Musculoskeletal (e.g. arthritis)	NSAIDs, prednisone, hydroxychloroquine, leflunomide, methotrexate, azathioprine
Renal	Cyclophosphamide, methylprednisolone, prednisone, mycophenolate mofetil, azathioprine
Serositis	NSAIDs, prednisone, hydroxychloroquine (flare preventions), other steroid sparing agents
Neuropsychiatric	Methylprednisolone, prednisone, cyclophosphamide, other steroid-sparing agents
Antiphospholipid antibody	Anti-coagulants
Hemolytic anemia, thrombocytopenia	Prednisone, immunoglobulin, azathioprine, and other steroid-sparing agents

Table 2.4: *Examples of treatments for various SLE symptoms.*

There appears to be a great variation in the use of different drug treatments among SLE patients. A study by Bernatsky *et al.* (2011) found 66% of their patients took antimalarial agents (hydroxychloroquine and chloroquine), 42.6% took prednisone, 16.6% took azathioprine, 7.5% took mycophenolate, 6.6% took methotrexate and 1.1% took cyclophosphamide over a 3-year period[79]. The study found that prednisone use (RR = 1.13, 95% CI: 1.09–1.19) and any immunosuppressant use (RR = 1.07, 95% CI: 1.01–1.14) were more often present in patients with higher damage than lower damage, as measured by the SLICC/ACR damage index (SDI). Increased disease activity as measured by the SLE Disease Activity Index (SLEDAI) 2000 was also associated with

prednisone use (RR = 1.04, 95% CI: 1.03–1.06) and any immunosuppressant use (RR = 1.05, 95% CI: 1.03–1.07); however, there was neither an association of higher SLE damage nor of higher SLE activity with antimalarial use. Two explanations were provided: (i) antimalarials are universally prescribed for SLE regardless of clinical severity and (ii) only when SLE has high activity are more aggressive immunosuppressants used.

SLE morbidity and mortality

SLE patients have been shown to have a greater risk of mortality than the general population with studies estimating the standardized mortality ratio (SMR) for SLE to be from 1.3–4.9[17, 27, 80, 81]. However, SLE survival has improved in recent years due to earlier disease recognition resulting in more timely and effective treatment[78, 80, 82]. The cause of death in SLE patients has been shown to change depending on SLE progression. SLE has been shown to have a bimodal pattern of mortality, where the landmark study by Urowitz *et al.* (1976) showed early deaths within one year of diagnosis were attributable to active SLE and infections. However, deaths in SLE patients after the first year had other causes, like myocardial infarction brought on by atherosclerosis[83]. Interestingly, most of these late deaths occurred while SLE was inactive. This bimodal pattern of mortality is still observed presently, but due to better disease management, this pattern has stretched out over a longer period of time[84] . While SLE survival has improved in recent years, research is still being conducted to determine SLE survival risk factors.

Risk factors for SLE survival

Chronic damage

Chronic damage is a major problem in SLE patients because this damage is irreversible and accumulates over the patients' lifetime, especially when not treated initially. Early damage is particularly devastating with one study finding a higher pulmonary SDI score predictive of death within ten years[85]. The study also found a higher renal SDI score was predictive of end-stage renal failure within ten years. Similar results were seen in another study where individuals who had early damage (score of at least one on the SDI) during their initial assessment had a 75% 10-year survival rate compared to a 92.7% 10-year survival rate in those who had no initial damage (score of zero)[86]. Predictors of higher SLE damage in multivariable models from different studies have included: African-American race/ethnicity, longer disease duration, prior damage, higher disease activity, older age at disease onset, lower educational level, treatment with corticosteroids (past and current), low income and exposure to cyclophosphamide at any time[70, 87, 88].

SLE disease activity

Several studies have associated SLE disease activity with higher mortality. In the LUMINA cohort, a higher Systemic Lupus Activity Measure score was associated with increased mortality (OR = 1.09, 95% CI: 1.01–1.17). Abu-Shakra *et al.* (1995) also found active SLE (SLEDAI score of at least 20) at presentation was associated with a two-fold risk of mortality (95% CI: 1.34–3.04)[89]. Having certain manifestations like thrombocytopenia[90], nephritis[89, 91], lung involvement[89], serositis[92] and neuropsychiatric manifestations[91] have been associated with lower survival. Disease activity itself is affected by factors such as race/ethnicity (e.g. with higher activity among African-Americans), younger age, lack of health insurance, past disease activity, available social support and abnormal illness-related coping behaviors[93, 94].

Age of SLE onset

Age of SLE onset is another risk factor. Jacobsen *et al.* (1999) found the standardized mortality ratio (SMR) of pediatric SLE patients was 67 (SMR = 67, 95% CI: 36–114)[27]. Abu-Shakra *et al.* (1995) also found high mortality ratio (SMR = 14.9, 95% CI: 6.0–30.7) in their pediatric SLE group (ages 0-24)[80]. While individuals with late-onset SLE present with benign symptoms, they have also been associated with lower survival[82, 90, 95]. However, this in large part reflects the higher mortality rate that comes with aging, even in the general population.

Sex

While SLE is predominantly found in females, it has been suggested that males face lower survival[82, 91, 95]. However, again, much of this effect reflects the higher mortality rate among men versus women, even in the general population.

Socioeconomic status

Socioeconomic status such as income, insurance status and education have been associated with mortality. One study found patients with an annual family income < $25,000 had poorer survival than those with incomes of ≥ $25,000 (HR = 1.7, 95% CI: 1.1–2.7)[82]. Another study found that for every $1,000 increase in annual household income, over-all mortality in SLE patients was reduced by 11% (HR = 0.89, 95% CI: 0.83–0.94)[95]. In the LUMINA cohort, being below the poverty line was associated with a four-fold increase in mortality (OR = 4.06, 95% CI: 1.50–11.01). Insurance status has been shown to be protective with one study finding a 34% reduction in SLE mortality when they had private medical insurance instead of no insurance (HR = 0.66, 95% CI: 0.46–0.96)[95]. Educational level has also been associated with mortality with one study finding

SLE-attributable deaths in Caucasians were significantly higher in individuals with lower education compared to those with higher education in both sexes[96]. However, this finding was not seen in African Americans or Asian women, which the authors attributed to SLE-related deaths being underreported in these two groups.

Race/ethnicity

Race/ethnicity is a possible mortality risk factor. Reveille *et al.* (1990) found African Americans had a lower survival compared to Caucasians[90]; however, the only socioeconomic factor adjusted for was insurance status, and it is possible that lower survival associated with race/ethnicity may have been confounded by income. In a study by Ward *et al.* (1995), no association between mortality at race/ethnicity was found after adjusting for medical insurance and census tract income level data[95].

SLE-related co-morbidities

Common SLE co-morbidities such as cardiovascular damage, osteoporosis (and osteopenia), infection and cancer affect quality of life and survival. Urowitz *et al.* (1976) and Nossent *et al.* (2007), showed deaths in surviving SLE patients are often caused by accumulated cardiovascular problems[83, 84]. A study by Manzi *et al.* (1997) showed that women with SLE compared to women from the Framingham Offspring Study were more likely to have a myocardial infarction for all age groups[97].

Lower bone mineral density (BMD) is more commonly observed in SLE patients compared to healthy individuals, putting them at risk for osteoporotic fractures[98]. One study found 50.8% of an SLE population had reduced BMD, 9.1% had at least one fragility fracture and 10.3% had osteoporosis[99]. Another study found 12% of women

self-reported having at least one fracture after SLE was diagnosed, and this fracture rate was 4.7-fold higher compared to the rate seen in the American women population[100].

Infection is one of the largest causes of morbidity and mortality[83, 84]. From 1990–2000 in the Euro-Lupus Cohort, 36% of SLE patients were diagnosed with an infection, and of those infected, 4.7% died[101]. It has been estimated that SLE patients are 5 and 2.6 times more likely to die from an infection (SMR = 5.0, 95% CI: 3.7–6.7) and pneumonia (SMR = 2.6, 95% CI: 1.6–4.1) than the rest of the population, respectively[81].

SLE patients have a greater risk of developing certain cancers like non-Hodgkin's lymphoma (NHL). Mellemkjer *et al.* (1997) found a 5.2-fold risk in SLE patients for developing NHL compared to healthy patients (standardized incidence ratio (SIR) = 5.2, 95% CI: 2.2–10.3), as well as an increased risk for lung, liver, vaginal and vulvar cancer[102]. SLE patients also have a higher risk of death from certain cancers. While one multisite, international cohort study found the SMR for any cancer to be 0.8 (SMR = 0.8, 95% CI: 0.6–1.0)[81], the same study reported SLE patients were more likely to die from three types of cancers: any hematologic cancer (SMR = 2.1, 95%: 1.2–3.4), non-Hodgkin's lymphoma (SMR = 2.8, 95% CI: 1.2–5.6) and lung cancer (SMR = 2.3, 95% CI: 1.6–3.0)[81].

As shown in the first part of this literature review, much research has gone into studying SLE. One data source that will help the epidemiological study of SLE is health administrative databases, discussed next.

Health administrative databases

Health administrative databases are an invaluable source of information on health care use by the population. In Quebec, there are four such databases: the

registration database, the physician claims (or physician services) database, the hospital discharge abstract database and the prescription drug database. These databases are managed by RAMQ, the Quebec's provincial health insurance agency. The registration database contains demographic information like: sex, age and date of death. The physician services database contains data surrounding physician visits, including: the physician seen, his or her specialty, the clinical setting and location of the visit, the reason for the visit as an International Classification of Diseases (ICD) code, the service provided and the date of the visit. The hospital discharge abstract database includes: sixteen discharge diagnosis codes (one primary and fifteen secondary discharge fields), the entry and discharge date from the hospital, treatments performed in the hospital and the date of treatment. The prescription drug database contains information for individuals covered under any of Quebec's prescription drug plans and contains the drug dispensed, the type of drug, the dosage and quantity prescribed, the pharmacist prescribing the drug and the dispensing date. All these databases can be linked together using a unique patient identifier, which provides researchers with a powerful tool to study diseases.

Past Quebec studies have used these databases for disease prevalence and incidence estimation[6]; health services use description[103, 104]; risk assessment studies with cohort[105] and case-control[105] designs; quality of care evaluation[106]; economic evaluation[107] and health policy evaluation[108]. One advantage to using these databases is they cover the majority of the population, which provides large sample sizes that facilitates the study of rare diseases such as SLE, and increases the generalizability of any results found[109].

Another benefit of these databases is their frequent update of records and their relatively low cost compared to surveys and other types of prospective data collection[109]. Administrative databases also do not suffer from participation bias, recall bias, non-response, or loss-to-follow-up (aside from patients that move out-of-province) [110]. These aspects make administrative databases an appealing resource to use in research and disease surveillance. However, like any database, health administrative databases are not without their biases and limitations, which are discussed next[36, 109].

Potential biases

Many factors influence what data are entered into health administrative databases, which can result in biases. Terris *et al.* (2007) addressed some of these influencing factors with his conceptual framework (Figure 2.1)[36].

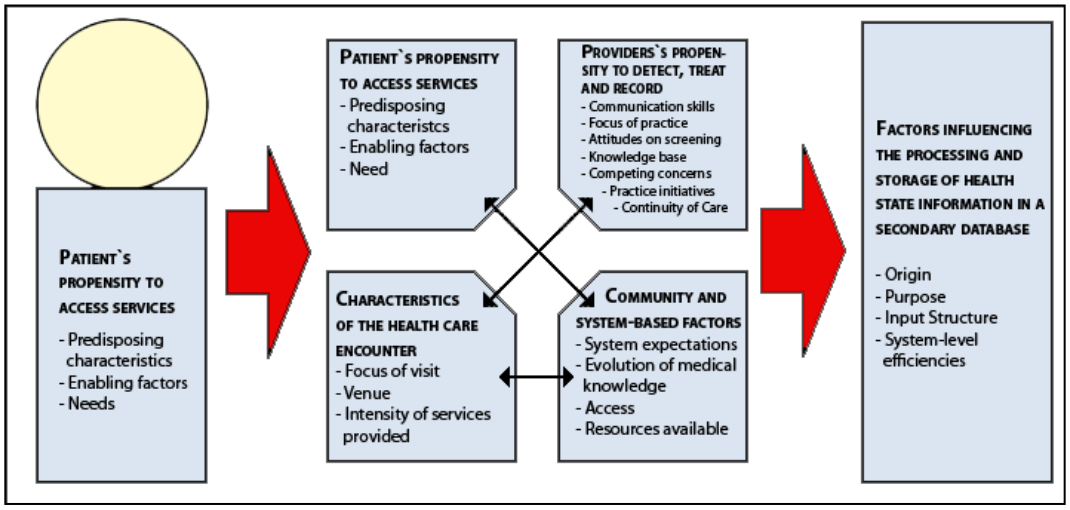

Figure 2.1: *Representation of Terris' conceptual framework* This figure is reconstructed from Figure 1 of Terris *et al.'s* (2007) 'Health state information derived from secondary databases is affected by multiple sources of bias'.

In his model, Terris suggests there are five main factors that impact an individual's health care encounter: (i) the patient's propensity to access services; (ii) the patient's propensity to report and adhere; (iii) the provider's propensity to detect, treat

and record; (iv) the characteristics of the health care encounter; and (v) community and system-based factors. These five factors, both independently and jointly, affect the proceedings of each health care encounter. The outcome of this health care encounter is then captured in physician billing information, which will become added to the physician claims database. The outcome is also captured in medical records, which are used by hospital clerks to code data into hospital discharge databases.

In addition to these five factors, Terris also identified two filters that affect the coding of this data. The first filter selects what information will be collected and stored, based on decisions by policy makers and the database creators. For example, in Quebec, clinical confirmatory data like laboratory test results are excluded.

The prescription drug database was designed to contain detailed drug data, but it still has limitations. Drug data pertaining to inpatients or individuals who reside in governmental long-term facilities is neither captured in the prescription drug database nor in the hospital discharge database[103], and this coverage gap can create bias in drug use studies where inpatient drug use is high or an elderly study population is considered[108]. As well, in Quebec, the RAMQ only covers select populations, including all residents aged 65 years and older, those on welfare and those without private drug insurance through their employment (including self-employed individuals and those unemployed[111]. Finally, this database only has information on outpatient drugs dispensed; it does not tell whether an individual actually complies with the prescription.

The second filter is created from the policies and procedures used for data coding. While strict coding procedures ensures consistency in hospitalization discharge databases, there is the risk that systematic recording could cause a misclassification

bias[109]. In contrast, physician claims databases have a wide variability in codes because coding is done directly by physicians, and their coding is not scrutinized; instead, it can be dictated by personal billing habits. As well, in Quebec, there are no incentives for physicians to include diagnostic information correctly on the claim, which would also increase variability of the data[112]. While most physicians still provide fee-for-service claims, the services of physician who are salaried are only available in administrative data if 'shadow' billing claims are required; however, in Quebec, this is not the case. Finally, it is also possible for physicians to bill patients directly for services instead of RAMQ, but this is rare in Quebec as only 1.4% of all physicians have opted out of the public system[113, 114].

The quality of data can vary as well because certain data are screened more carefully. For example, procedure codes are audited more carefully than diagnostic information[110]. Errors can also occur; in Quebec, one study of the prescription drug database found the patient's Medicare number to be incorrect 0.4% of the time, compared to a 30.9% and 27.9% error in the quantity and duration of the drug prescribed[103].

Finally, coding practices change over time. Many databases continue to use the ICD 9th revision (ICD-9), but many others have switched to ICD-10 because it contains more numerous codes[115]. In Quebec, this switch occurred in Med-Echo data in April 2006, but RAMQ physician billing data continue to use ICD-9 codes. There are potential issues of comparability between revisions, but at least one study has found that the two coding systems produce comparable Charlson comorbidity prevalence estimates, with the ICD-10 codes performing slightly better[115] .

Terris's five factors, which are all upstream of data entry, impact what these databases ultimately capture, and these biases and errors can manifest themselves in different ways. When administrative databases are used for ascertaining diseases, the data may be correct, but they may represent inaccurate diagnoses, premature diagnoses, missing diagnoses or coding errors[116]. These problems complicate the use of these databases, but strategies have also been utilized to address these problems.

Case ascertainment limitations

Health administrative databases were originally designed for administrative and planning purposes, not for research[36, 109, 117]. The two databases most often used for disease case ascertainment are the hospitalization discharge abstract database and the physician claims database. For both these databases, the first and foremost limitation is that the diagnostic codes are not necessarily clinically confirmed, and there is no method to confirm disease cases without medical chart review, which itself may be fraught with difficulties. A diagnosis code could mean a disease case, but it may also represent a working diagnosis, an incorrect diagnosis or a coding error. This issue may be particularly important for clinically complex diseases like SLE where there is no one definitive diagnostic test.

A second limitation is these databases have a limited number of disease code fields, which means the absence of a diagnosis code does not necessarily mean the absence of the disease in the individual[118]. In Quebec, the hospitalization discharge abstract database has sixteen fields, but (as in most other provinces) the physician claims database allows only one diagnosis per physician visit. Hence, a person with

numerous medical conditions cannot have all the diagnoses recorded at a single visit. Database-specific limitations also exist.

Physician claims database limitations

As mentioned earlier, the Quebec physician claims database only records one diagnostic code per visit, which means for a patient with multiple diseases, a physician must choose a code. For example, when an SLE patient visits a nephrologist for treatment of nephritis, the nephrologist can choose to code for either the manifestation or for SLE. The physician might be more inclined to bill for nephritis because it is the cause of the visit, but he or she might also bill for SLE, the underlying cause of the nephritis. Missing data is also an issue because missing diagnostic codes decreases sensitivity. A study by Wilchesky *et al.* (2004) found about 30% of Quebec physician claims contained either missing or unusable diagnostic codes compared to general practitioner medical records[110].

Hospitalization discharge abstract limitations

Ascertained cases from hospitalization data are considered to be more accurate, the main reason being that because more severe cases of a disease will require hospitalization, the accompanying diagnosis codes will intrinsically be more accurate since a severe disease case is more identifiable[119]. However, if only this database is used, the improved sensitivity comes at the cost of selecting only a subgroup of patients with severe disease. And, of course, hospitalization data alone does not allow researchers to study disease progression and health services use.

Methods for addressing limitations

Many methods have been used to address the limitations inherent in administrative data. Some have attempted to validate cases found using another source as a reference standard, which is most commonly medical records[116, 120-125]. Other reference standards have included diagnostic tests[126], patient self-report[120, 124, 127, 128] and previously validated case ascertainment algorithms[129]. The basic concept is to compare the cases detected by a case ascertainment algorithm to a standard like medical records, calculating sensitivity and specificity, similar to the validation of a new diagnostic test.

Many case ascertainment algorithms have been used in the past, and the simplest algorithm for case detection is the presence of one corresponding diagnostic code. However, this rule can lead to a high false positive rate, so many algorithms require multiple claims[116, 120, 121, 124]. Even with a multiple claims definition, there may still be false positives, particularly if a cluster of billing claims represents a working diagnosis that is later wrong. To address this issue, algorithms have also required a condition as to how the multiple billing codes occur together. Two commonly-used conditions are the diagnoses codes must occur within a maximum time period or after a minimum waiting period[6, 129-132]. For hospitalization data, one code is often considered sufficient because hospitalization diagnoses, being abstracted by trained medical clerks according to a specific protocol, are considered to be more accurate[6].

Knowing the sensitivity and specificity of case detection algorithms is important so that adjustments can be made to obtain unbiased estimates[109]. The importance of adjustment was emphasized by Ladouceur *et al.* (2007) who showed that while health administrative databases contain minimal selection bias, not adjusting for case

misclassification can still lead to biased estimates with confidence intervals that are too narrow[133]. This is often not a major problem because many algorithms feature a high specificity, though of course, at the cost of lower sensitivity[6, 110, 120, 123, 129].

Although medical records are commonly used as a gold standard, there may be difficulties with accessing medical charts, or the level of detail documented within the charts may be insufficient to determine whether or not an individual meets specific diagnostic criteria. In the situation where a gold standard is absent or not feasible, there are various options. One strategy is to use latent class methods, where case status is not specifically observed; rather, a latent class regression model relates a set of observed multivariable variables to the set of latent variables. For case ascertainment, the specificities and sensitivities of the definitions are not known initially, but they can be estimated by using information that is contributed by each case definition's ability to identify the true, but *latent* (or unknown) disease status of an individual. Hierarchical elements can be added to the model to allow variation at multiple levels such as sex, age, and rural/urban residency and physician characteristics[6].

SLE-specific validation studies

Several SLE-specific validation studies have been performed. The first validation study was by Katz *et al.* (1997), and they estimated the sensitivity and positive predictive value (PPV) of SLE using Medicare Part B physician claims data[134]. Using a single diagnosis code definition, they calculated a sensitivity of 0.85 (95% CI: 0.73–0.97) and a PPV of 0.90 (95% CI: 0.81–0.99) using medical abstracts as the standard. A study by Lim *et al.* (2007) calculated the sensitivity and specificity of an SLE case definition based on ICD-9 codes from a hospital discharge database for a one-year period and a 5-year

period[135]. For a one-year period, SLE sensitivity was 67% and specificity was 93.5%, and for a 5-year period, sensitivity was 79.3% and 90.2%.

Bernatsky *et al.* (2007) estimated the sensitivity and specificity for a physician claims algorithm and a hospitalization discharge algorithm using a Bayesian latent class hierarchical model[6]. Individuals were identified as having SLE from physician billing claims if they had two ICD-9 SLE-coded (710.0) physician visits within two years but at least two months apart. There was no gold standard used in this study, so an individual's true disease status was estimated using a latent class model. The sensitivity and specificity estimates were 44.8% (95% CI: 43.1–46.6) and 99.99% (95% CI: 99.99–100). The hospitalization algorithm was at least one SLE-coded hospital discharge, and the sensitivity and specificity were 58.4% (95% CI: 56.2–60.5) and 99.99% (95% CI: 99.99–100). A study by Chibnik *et al*. (2010) used medical records to validate algorithms focused only on lupus nephritis (ICD-9: 580.00-586.00 and 791.0). They studied four algorithm definitions: (i) at least three lupus nephritis diagnosis codes, (ii) at least three nephrologist visits, (iii) either of the two definitions and (iv) both of the definitions[119]. For these algorithms, sensitivity was not given, but PPV estimates varied from 89% to 92% for identifying SLE, and between 79% and 88% for identifying lupus nephritis.

Bernatsky *et al.* (2011) recently assessed administrative data-based definitions for a number of systemic autoimmune rheumatic diseases (SARDs), including SLE[125]. Using three algorithms – at least one SLE-coded hospitalization discharge diagnosis, at least two SLE physician billing code diagnoses at least 2 months apart or at least one rheumatologist billing code SLE – the sensitivity and specificity of the full algorithm

compared to chart review (the gold standard) were 98.2% (95% CI: 95.5–99.3) and 72.5% (95% CI: 68.7–75.9).

Conclusions

SLE is a complex disease. The yearly incidence has been estimated to be between 0.1 and 18.5/100,000 and prevalence estimates to lie between 20 and 70/100,000. Known risk factors include: sex, Epstein-Barr virus, age, genetics, smoking, and occupational factors (silica and UV radiation exposures). Individuals experience different manifestations and varying levels of disease severity, and unfortunately, SLE patients have lower survival compared to the population. Research has shifted away from estimating the burden of disease in the population towards understanding risks for SLE morbidity and mortality. To help study SLE longitudinally, health administrative databases are a potential source. These databases are advantageous because they cover virtually the entire population which reduces selection bias and increases the generalizability of results. Administrative databases also have their limitations, including those that pertain to ascertaining disease cases. Only a few studies have estimated SLE case ascertainment algorithm sensitivities and specificities, and these are difficult to compare, because of differences in case definition, data sources, and analytic methods. To further examine this matter, in the next chapter (Chapter 3), we will examine how applying different SLE case ascertainment algorithms can change incidence and prevalence rates, even when using the same data sources.

Chapter 3: Ascertainment algorithms for systemic lupus erythematosus incidence and prevalence estimation from Quebec administrative data

Authors: R. Ng, S. Bernatsky, E. Rahme

Abstract

Background: Ascertaining disease cases using health administrative databases is a current methodological challenge. This study examined systemic lupus erythematosus (SLE) incidence and prevalence in Quebec from 1989-2003 using various case ascertainment algorithms and data period lengths.

Methods: An initial SLE case definition was based on a previously-used, SLE algorithm with three criteria: one SLE hospital discharge code, one rheumatologist SLE claim and/or two SLE non-rheumatologist claims at least eight weeks apart but within two years. Case ascertainment was performed using alternative algorithms with different definitions and different data period lengths. The SLE incidence and prevalence estimates based on the different approaches to case ascertainment were calculated and compared to the estimates from the initial SLE case definition.

Results: The 1998 SLE incidence based on the initial SLE case definition was 6.0 cases per 100,000 individuals (/100,000) (95% confidence interval (CI), 5.5–6.6), the 2001 incidence was 5.7/100,000 (95% CI: 5.2–6.3) and the prevalence was 65.5/100,000 (95% CI: 63.7–67.4). When the initial algorithm was modified by changing one of its parameters, the 1998 incidence ranged from 4.4–7.4/100,000 and the prevalence ranged from 47.8–79.1/100,000. When the length of the data period for case

ascertainment was shortened by one-year increments, from fifteen years to three years, prevalence estimates decreased and incidence estimates increased. When SLE cases were ascertained over a 10-year data period, the prevalence (61.8 cases per 100,000) was only 4.1% less than the prevalence estimate generated from the full 15-year data period (65.5/100,000). Using this 10-year period, the SLE incidence estimate in 2001 (6.1/100,000) was only 7.3% higher than the initial 2001 incidence estimate (of 5.7/100,000). However, when the 5-year data period was used instead of the 15-year data period SLE cohort, prevalence decreased by 29.9% to 46.0/100,000 (95% CI: 44.4–47.5), and incidence was falsely inflated by 38.3% to 7.9/100,000 (95% CI: 7.3–8.5).

Conclusion: This study showed that varying the algorithm definition and the data period can change incidence and prevalence estimates considerably, so all algorithm features, including the length of time in which the data spans, should be selected carefully and explicitly stated.

Introduction

Health administrative databases are increasingly being used to study health issues such as incidence and prevalence of diseases[6], health services use[136] and medication prescription patterns[103]. Two comprehensive health administrative databases in Canada include hospital discharge databases which summarize inpatient visits, and physician claims (billing) databases which track physician reimbursement for inpatient and outpatient visits (drug data is not comprehensively available in all provinces). In single-payer systems like Canadian provinces, these databases are attractive for research purposes because they cover virtually the entire population, which facilitates the study of rare events and the generalizability of study results[109].

However, these databases were designed for managing reimbursement, not research, so there are data limitations. One important limitation is an inability to confirm an individual's true disease status because billing and hospital diagnostic codes are not necessarily clinically confirmed.

Most times, when administrative data are used for surveillance and/or research, cases are ascertained with algorithms that use data fields such as diagnosis codes[6, 110, 120, 122-124, 129], procedural codes[122] and/or dispensed drugs[124] to infer disease status. For a handful of diseases, the quality of these algorithms has been validated against a reference standard (often medical records) to determine sensitivity and specificity[1, 5-10]. Different algorithms have been used (Table 3.1), but they often feature similar elements [6, 110, 120, 122-125, 129]. To ascertain cases, these algorithms are applied within a specified data period to select individuals who fit the case definition. Some criteria require the appearance of only one code; for example, many algorithms require only one hospital discharge diagnosis code because each discharge is a summary of the entire inpatient stay. However, other criteria may require multiple, conditional diagnostic codes; for example, a criterion requiring two or more physician claims over a defined interval. This is because the accuracy of physician claims diagnostic codes are subject to more variability due to physicians' personal billing habits[137], and since diagnostic codes are not audited for accuracy.

Author(s)	Diseases studied	Algorithm
Bernstein *et al.* (1999)	Inflammatory bowel disease (IBD): Crohn's disease, Ulcerative colitis	(i) For individuals residing in the province for at least 2 years: at least 5 separate physician claims and/or hospitalizations for IBD; (ii) For individuals residing in the province less than 2 years: at least 3 separate physician claims and/or hospitalizations for IBD
Hux *et al.* (2002)	Diabetes mellitus (DM)	(i) At least 1 ICD-9 DM physician claims or at least 1 hospital discharge claim (ii) At least 2 ICD-9 DM physician claims or at least 1 hospital discharge claim
Wilchesky *et al.* (2004)	Hypertension, peptic ulcer disease, diabetes, chronic obstructive pulmonary disease, prostatic hypertrophy, asthma, congestive heart failure, gout, renal failure, glaucoma, postural hypotension, heart block and Raynaud's disease	At least 1 disease-specific physician billing claim
Bernatsky *et al.* (2007)	Systemic lupus erythematosus (SLE)	(i) at least 1 ICD-9 SLE hospital discharge diagnosis, or (ii) at least 2 ICD-9 (710.1) physician claims at least 2 months apart but within 2 years
Ladouceur *et al.* (2007)	Osteoarthritis (OA)	Multiple definitions tests: (i) 1 OA visit in 2002 (ii) 1 prescription for acetaminophen or non-steroidal anti-inflammatory drug , but not methotrexate or plaquenil (iii) Injection procedure common to OA patients, arthroplast or tibial osteotomy in 2002
Lacaille *et al.* (2008)	Rheumatoid arthritis (RA)	At least 2 RA physician visits more than 2 months apart while excluding those with either 2 visits subsequent to the second RA visit for other inflammatory arthritides or did not have a confirmatory RA diagnosis by a rheumatologist if a rheumatologist was seen
Guttmann *et al.* (2010)	Diabetes mellitus (DM)	Multiple definitions tested: (i) 1 - 5 DM claims in 1 year (ii) 1 - 5 DM claims in 2 years (iii) 1 - 3 DM claims in 1 year or 1 DM hospital record (iv) 1 -3 DM claims in 2 years or 1 DM hospital record
Marrie *et al.* (2010)	Multiple sclerosis (MS)	(i) At least 7 separate physician claims, hospitalizations or prescription claims for MS from 1984 to 1997, or (ii) At least 3 separate physician claims, hospitalizations or prescription claims for MS between 1998 and 2006
Bernatsky *et al.* (2011)	Systemic lupus erythematosus (SLE)	(i) At least 2 ICD-9 SLE physician claims at least 2 months apart (ii) at least 1 ICD-9 SLE claim by a rheumatologist or internist, or (iii) at least one ICD-9 SLE hospital discharge diagnosis

Table 3.1: *Examples of previously-used case ascertainment algorithms within Canadian health administrative databases.*

The data period used for case detection is also important. While health administrative databases can have wide population coverage, they rely on patient-initiated contact with the health system, so patients with mild and/or inactive chronic disease cases (or with less access to care) will be under-detected. As well, individuals

with multiple chronic diseases may have one or more of these conditions underreported by their general practitioner because most provincial databases (including Quebec) allow physicians to bill only a single diagnostic code per patient visit. When a researcher obtains more years of administrative data, this longer data period should increase case ascertainment sensitivity, and should increase the ability to distinguish between incident and prevalent cases. However, obtaining longer data periods may be unfeasible, or very costly. In addition, it adds potential time-related variation regarding data quality, and/or trends in physician services use and billing. Thus, the data period for ascertaining cases of chronic rheumatic diseases is important, but to our knowledge, quantitative data period effects of any chronic disease have not been documented extensively with Canadian administrative data.

For this study, we focused on systemic lupus erythematosus (SLE), a relatively uncommon disease with previously-used ascertainment algorithms. Previous studies with varying methodological approaches have estimated SLE yearly incidence to be 1.0–7.6/100,000 and prevalence to be 19.3–94.0/100,000[2]. SLE affects mainly females, with about 90% of adult clinical cohorts being women. SLE is a chronic disease with a relapsing-remitting nature, which means SLE may go undetected in administrative data, especially within shorter data periods.

Objectives

The objective was to determine the effects of various claims-based case ascertainment algorithms on incidence and prevalence estimates, as well as the impact of data period length over which to detect SLE cases.

Methods

Health administrative databases

The databases used were the Québec Maintenance et exploitation des données pour l'étude de la clientèle hospitalière (Med-Echo) hospitalization discharge abstract database, the Régie de l'assurance maladie du Québec (RAMQ) physician claims database and the RAMQ registration database. The hospitalization database contains sixteen discharge diagnosis fields, one primary diagnosis code, and up to fifteen secondary diagnosis codes. The physician claims database contains one diagnosis code per patient visit. Both databases use diagnosis codes that follow the International Classification of Disease 9th revision (ICD-9) classification system. The registration database contains demographic data such as sex, date of birth and date of death. Data for this study spanned a 15-year period from January 1, 1989 to December 31, 2003.

SLE ascertainment algorithms and cohort formation

Six algorithm types were used as SLE case definitions (Table 3.2): an initial algorithm and five sets of alternative algorithms (*single-rule algorithms; modified single-rule algorithms*; *1-modified multi-rule algorithms*; *2-modified multi-rule algorithms*; and *data period variation algorithms*).

Type of Algorithm	Example of Algorithm	The difference in the definition compared to the initial algorithm
Initial algorithm	In a 15-year data period: (i) 1 SLE hospital discharge (ii) 1 SLE rheumatologist claim, and/or (iii) 2 SLE physician claims 8 weeks apart but within 2 years	No difference
Single-rule algorithm	In a 15-year data period: 2 SLE physician claims 8 weeks apart but within 2 years	Missing the hospital and rheumatologist criteria
Modified single-rule algorithm	In a 15-year data period: 2 SLE physician claims 8 weeks apart but within (1 to 5) years	In a 15-year data period: 2 SLE physician claims 8 weeks apart but within 2 years; Also missing the hospitalization and rheumatologist criteria
1-modified multi-rule algorithm	In a 15-year data period: (i) 1 SLE hospital discharge (ii) 1 SLE rheumatologist claim and/or (iii) 2 SLE physician claims 8 weeks apart but within (1 to 5) years	In a 15-year data period: (i) 1 SLE hospital discharge (ii) 1 SLE rheumatologist claim, and/or (iii) 2 SLE physician claims 8 weeks apart but within 2 years
2-modified multi-rule alorithm	In a 15-year period: (i) 1 SLE hospital discharge (ii) 1 SLE rheumatologist claim and/or (iii) 2 SLE physician claims (1 day to 16 weeks) apart but within (1 to 5) years	In a 15-year data period: (i) 1 SLE hospital discharge (ii) 1 SLE rheumatologist claim, and/or (iii) 2 SLE physician claims 8 weeks apart but within 2 years
Data period variations	In a (3 to 14)-year data period: (i) 1 SLE hospital discharge (ii) 1 SLE rheumatologist claim, and/or (iii) 2 SLE physician claims 8 weeks apart but within 2 years	In a 15-year data period: (i) 1 SLE hospital discharge (ii) 1 SLE rheumatologist claim, and/or (iii) 2 SLE physician claims 8 weeks apart but within 2 years

Table 3.2: *Types of case ascertainment algorithms examined.* The underlined criteria identify the difference between the alternative algorithms and the initial algorithm.

Initial algorithm and the initial SLE case definition

A previously-used, multi-rule algorithm was employed as the *initial algorithm* to detect SLE cases over a 15-year data period[6, 125]. This initial algorithm consisted of the following three criteria:

1) One SLE-coded (ICD-9: 710.0) hospital discharge in the Med-echo database; and/or
2) One SLE-coded rheumatologist claim from the RAMQ database and/or
3) Two SLE-coded non-rheumatologist physician claims at least eight weeks apart but within two years of each other from the RAMQ database.

To be a case, at least one definition needed to be fulfilled.

Alternative algorithms and alternative SLE cohorts

The first set of alternative case definitions, *single-rule algorithms,* was based on one rule (*single-rule*), which was one of the three initial algorithm criteria.

Three additional sets of SLE case definitions were formed by using criteria that were modified from criteria (2) and (3). Criterion (2) was changed in two ways by:

a) Increasing the number of rheumatologist visits required from one visit to three visits; and
b) Expanding the definition from rheumatologists to include other specialists who tend to have greater SLE clinical experience (internists, nephrologists, and allergist/immunologists).

Criterion (3) was changed in three ways by:

c) Varying the minimum waiting period between the two physician claims from eight weeks to complete relaxation of the waiting period to sixteen weeks;
d) Shifting the maximum time period between the two physician claims visits from a two-year maximum to one-year and 5-year maximums; and
e) Removing the time period constraints between the two physician claims altogether, requiring instead, a total number of physician claims that varied from two to four visits that could occur at any time during the 15-year data period.

The second set of alternative SLE case definitions, *modified single-rule* case definitions, relied only on one rule (*single-rule*), but this rule was one of the five *modified* criteria from criteria 2 or 3 ((a) to (e)).

The third set of case definitions, *1-modified multi-rule algorithms*, was based on requiring all three initial (1 hospitalization, 1 rheumatologist and/or 2 non-rheumatologist diagnostic codes) criteria (*multi-rule*), but with one of the five alternative criteria replacing its corresponding initial criterion (*1-modified*).

The fourth set of case definitions, *2-modified multi-rule algorithms,* used all three initial algorithm criteria (*multi-rule*) as well, but with two alternative criteria replacing their corresponding initial criteria (2-*modified*) (Table 3.2). Some two alternative criteria combinations were not possible, so these combinations were excluded; for example, the alternative criteria that varied the minimum waiting period

between visits (criterion (c)) and the maximum time period between visits (criterion (d)) were not combined with criterion (e), which ignored the time period constraints altogether. As well, combinations of alternative criteria where one criterion definition was a subset of another were also excluded; for example, an algorithm that required three SLE rheumatologist visits (criterion (a)) was not used in combination with criterion (e), which required three SLE physician visits, because all cases detectable with criterion (a) were also detectable with criterion (e).

The final set of alternative SLE case definitions was derived using *data period variations* where the initial algorithm detected cases over twelve shorter, artificially created data periods (*data period variations*) that varied in length from fourteen years (1990-2003) to three years (2001-2003). Here, 'data period' is the period of time in which claims data are available for analysis. Shortening the data period results in less time available in which to detect SLE cases and distinguish between incident and prevalent cases. The twelve shorter data periods differed by one year intervals, and the one-year interval that was always excluded between successive data periods was the earliest year of data. By always excluding the earliest years, each shorter data period was progressively more 'left-censored' compared to the 15-year data period, which allowed for progressive comparisons of the results to the use of a 15-year data period.

Yearly incidence

Yearly incidence and its 95% CI were calculated by selecting all incident cases in a year and dividing them by the population of Quebec on July 1st of that year. To be an incident case in a particular year, the individual was required to have their index claim in that year. The index claim was selected as the earliest SLE-coded visit with no prior SLE

visits in a defined 'washout period'. Specific years for calculating yearly incidence were chosen to maximize the length of the washout period and to ensure all ascertainment algorithm criteria had the appropriate length of time to detect cases. Thus, when yearly incidence was calculated using the alternative algorithms, SLE incidence in 1998 was calculated because it was the first year in which a complete 5-year period was available (1999-2003), as required by alternative criterion (e), when up to five years between visits was allowed. Calculating incidence in 1998 also meant a 9-year (1989-1998) washout period was being used, the longest washout period possible for the data at hand. For these algorithms, the 2001 incidence was calculated because initial criterion (3) (two non-rheumatologist visits at least eight weeks apart but within two years) required a maximum of two years after the first SLE visit for the second visit to occur (2002–2003). For data periods of fifteen years (1989–2003), ten years (1994–2003), five years (1999–2003) and three years (2001–2003), the length of the washout periods decreased from twelve years, to seven, two and zero years. With a decreasing washout period length, there would be more *misclassified (as incident) prevalent SLE cases* (Figure 3.1). Misclassified (as incident) prevalent SLE cases were cases identified as a prevalent case using the 15-year data period, but falsely considered as incident cases within the shorter data period because the real SLE index claim did not occur within the shorter data period. For these cases, a later SLE claim observed in the shorter data period would be incorrectly identified as the index claim, and this false index claim would result in the prevalent case being misclassified as incident. Misclassified (as incident) prevalent SLE cases will falsely inflate incidence estimates.

Prevalence

SLE prevalence and its 95% CI was calculated by dividing all cohort members alive on December 31, 2003 by the population of Quebec on July 1st, 2003[138]. For the data period variation cohorts, shorter data periods meant there would be *undetected SLE cases* (Figure 3.1). Undetected SLE cases were cases ascertained within the 15-year data period and alive on December 31, 2003, but were not ascertained as SLE cases by the initial algorithm within the shorter data period because no initial algorithm criteria were met. Undetected SLE cases will underestimate prevalence figures.

Percent change

To compare the alternative algorithms and the data period variation case definitions to the initial algorithm, the percent change in incidence and prevalence estimates were calculated by dividing the difference of the algorithm or data period being examined and the initial, 15-year data period algorithm by the 15-year data period algorithm.

Statistical analyses

Incidence and prevalence estimates were compared based on the number of cases identified by each SLE algorithm. The 1998 incidence was calculated first using the initial algorithm. Then, estimates were generated using: the single-rule algorithms; the modified single-rule algorithms; the 1-modified multi-rule algorithms; and the 2-modified multi-rule algorithms. The 2001 incidence was also calculated for the full 15-year data period, and all shorter data periods. The period prevalence was calculated for all SLE algorithms. The percent change in incidence and prevalence estimates, relative to the initial SLE case definition, were calculated for all alternative case definitions. For

each data period variation, the number of misclassified (as incident) prevalent SLE cases and undetected SLE cases were tabulated.

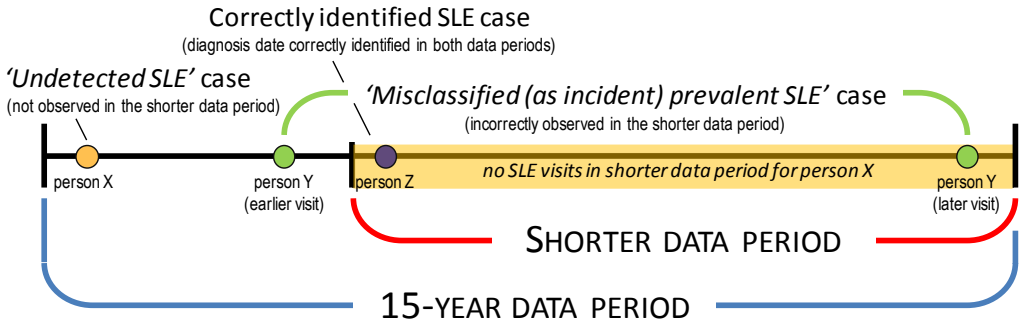

Figure 3.1: *Comparing the 15-year data period to a shorter data period with examples of correctly identified, misclassified (as incident) prevalent and undetected SLE cases.*

Results

Initial case definition using the 15-year data period

The initial algorithm, requiring the fulfillment of either: one SLE hospitalization, and/or one SLE claim by a rheumatologist, and/or two SLE physician claims at least eight weeks apart but within two years was first applied over the 15-year data period. A total of 5735 cases were detected for a prevalence of 65.5 cases per 100,000 individuals (95% CI: 63.7–67.4). The 1998 incidence was 6.0/100,000 (95% CI: 5.5–6.6). In 2001, the one-year incidence was 5.7/100,000 (95% CI: 5.2–6.3).

Alternative SLE algorithm comparisons

'Single-rule' algorithms

Across the three *single-rule* algorithms, the number of SLE cases detected varied between 2994 and 3213 cases, compared to the 5735 cases identified using the initial case definition. All three 1998 incidence and prevalence estimates based on single-rule algorithms were significantly different from the initial algorithm estimates (Table 3.3).

SLE case ascertainment algorithm	1998 incidence and 95% CI (/100,000)	Prevalence and 95% CI (per 100,000)
Initial algorithm	6.0 [5.5- 6.6]	65.5 [63.7- 67.4]
Single-rule algorithms		
(1) 1 hospitalization	2.6 [2.3- 3.0]	32.0 [30.7- 33.2]
(2) 1 rheumatologist visit	3.7 [3.2- 4.1]	38.3 [36.9- 39.7]
(3) 2 physician visits 8 weeks apart but within 2 years	3.1 [2.7- 3.5]	33.1 [31.8- 34.4]
Modified single-rule algorithms		
(a) Number of rheumatologist visits: (i) 3 visit	1.9 [1.6- 2.2]	18.1 [17.1- 19.0]
(b) Specialists: (ii) other specialists included	5.7 [5.1- 6.2]	58.2 [56.4- 59.9]
(c) Minimum waiting period between visits: (i) 16 weeks	2.8 [2.4- 3.1]	27.6 [26.5- 28.8]
(ii) no waiting period	3.7 [3.3- 4.2]	45.6 [44.0- 47.1]
(d) Maximum time period between visits: (i) 1 year	2.8 [2.4- 3.2]	30.4 [29.2- 31.7]
(ii) 5 years	3.3 [2.9- 3.7]	38.9 [37.4- 40.3]
(e) Number of physician visits: (i) 4 visits	2.7 [2.3- 3.0]	30.1 [28.9- 31.4]
(ii) 2 visits	4.4 [3.9- 4.8]	35.8 [34.4- 37.1]
1-modified multi-rule algorithms		
(a) Number of rheumatologist visits: (i) 3 visit	4.8 [4.3- 5.3]	52.1 [50.5- 53.7]
(b) Specialists: (ii) other specialists included	7.2 [6.6- 7.8]	76.0 [74.0- 78.0]
(c) Minimum waiting period between visits: (i) 16 weeks	5.9 [5.4- 6.5]	63.6 [61.8- 65.4]
(ii) no waiting period	6.2 [5.6- 6.8]	72.3 [70.4- 74.2]
(d) Maximum time period between visits: (i) 1 year	5.9 [5.4- 6.5]	64.7 [62.9- 66.5]
(ii) 5 years	6.2 [5.6- 6.8]	67.6 [65.7- 69.4]
(e) Number of physician visits: (i) 4 visits	6.1 [5.5- 6.7]	64.3 [62.5- 66.1]
(ii) 2 visits	6.7 [6.1- 7.3]	66.9 [65.0- 68.7]

Table 3.3: *Incidence and prevalence estimates for single-rule, modified single-rule and 1-modified multi-rule algorithms*. The roman numerals denote whether the minimum/maximum value was the most restrictive (denoted by (i)) or the least restrictive in detecting SLE cases for that alternative criteria (denoted by (ii)).

'Modified single-rule' algorithms

The number of cases in the *modified single-rule* algorithms varied from 1532 to 4908 cases. All prevalence and 1998 incidence estimates calculated from the minimum and maximum vales for each modified single-rule algorithm were significantly different than the initial cohort estimates, except for the incidence from modified single-rule algorithm criterion (b) (Table 3.3).

Incidence and prevalence estimates were also compared between the modified single-rule algorithms and the single-rule algorithms they were derived from. For incidence estimates, six of the eight estimates calculated from the minimum and

maximum values for each modified single-rule algorithm were significantly different than their single-rule counterparts. Similarly, all eight modified single-rule prevalence estimates were significantly different than their single-rule counterparts. The modified single-rule algorithm requiring three SLE rheumatologist claims (criterion (a)) showed the greatest decrease in incidence (69.1%) and prevalence (72.4%), compared to our initial estimates (i.e. incidence of 1.9/100,000 and prevalence of 18.1/100,000). The modified single-rule algorithm which included other specialists with SLE clinical experience (criterion (b)) was associated with a non-significant decrease in 1998 incidence and a very small statistically-significant decrease in prevalence.

'1-modified multi-rule' algorithms: effects due to altering one item at a time

With the multi-rule algorithms (hospitalization, rheumatology and physician definitions) where one item was varied at a time, the alternations led to variations in the total number of cases ranging from 4639 to 6602 (prevalence and incidence estimates are shown in Table 3.3). Of the eight 1998 incidence estimates determined, only two were statistically different from the initial estimates. Similarly, of the eight different 1998 prevalence estimates, only three were statistically different from the initial estimates. The largest decrease in incidence and prevalence (compared to the initial algorithm) was seen when the alternative criterion (a) (three rheumatologist SLE visits) was used (both by 20.5%). Including other specialist diagnoses in addition to rheumatologist diagnoses (criterion (b)) resulted in higher incidence and prevalence estimates compared to the initial algorithm, by 18.9% and 15.9%, respectively.

Two alternative criteria changes			(a) Rheum / specialist visits	(b) Specialist expansion	(d) Maximum time between physician visits	
			(i) 3 visits		(i) 1 year	(ii) 5 years
(a) Rheumatologist / specialist visits	(i) 3 visits	Prev		53.5 [51.9- 55.2]		
		Inc		5.0 [4.5- 5.5]		
(c) Minimum waiting period	(i) 16 weeks	Prev	50.5 [48.8- 52.1]	75.6 [73.6- 77.6]	63.4 [61.6- 65.2]	66.2 [64.3- 68.0]
		Inc	4.6 [4.2- 5.1]	7.1 [6.5- 7.7]	5.8 [5.3- 6.4]	6.1 [5.5- 6.7]
	(ii) no period	Prev	55.1 [53.4- 56.8]	77.0 [75.0- 79.0]	66.5 [64.7- 68.3]	68.8 [66.9- 70.6]
		Inc	4.9 [4.4- 5.4]	7.2 [6.6- 7.8]	6.1 [5.5- 6.6]	6.3 [5.7- 6.9]
(d) Maximum time between physician visits	(i) 1 year	Prev	47.8 [46.2- 49.3]	75.2 [73.2- 77.1]		
		Inc	4.6 [4.1- 5.1]	7.1 [6.5- 7.7]		
	(ii) 5 years	Prev	60.7 [59.0- 62.5]	79.1 [77.1- 81.1]		
		Inc	5.0 [4.5- 5.5]	7.2 [6.6- 7.8]		
(e) Physician visits	(i) 4 visits	Prev	50.3 [48.7- 51.9]	75.7 [73.7- 77.6]		
		Inc	4.4 [3.9- 4.9]	7.0 [6.4- 7.6]		
	(ii) 2 visits	Prev	53.9 [52.2- 55.5]	76.3 [74.3- 78.2]		
		Inc	5.7 [5.1- 6.2]	7.4 [6.8- 8.0]		

Figure 3.2: *SLE incidence and prevalence for two simultaneous criteria changes.*

'2-modified multi-rule' algorithms: variations in two items at a time

With the multi-rule algorithm (hospitalization, rheumatology and physician definitions) where two items were varied at once, the number of SLE cases detected varied from 4294 to 6861. As expected, the algorithm that detected the fewest cases was the most stringent definition, and it had a definition of: one SLE hospitalization discharge, three SLE rheumatologist visits and four SLE physician claims. The 1998 incidence and prevalence with this definition were respectively 27.0% and 23.3% smaller than the initial estimates (incidence of 4.4/100,000 and prevalence of 50.3/100,000; see Figure 3.2). The algorithm that detected the most cases had a broader definition, consisting of: one SLE hospitalization discharge, one SLE visit to any of the defined specialists experienced with treating SLE and two SLE physician visits at any time. The 1998 incidence and prevalence estimates with this definition were respectively 22.5% and 16.4% larger than the initial estimates (incidence of 7.4/100,000 and prevalence of 76.3/100,000).

Data period variations

With the various data periods, the number of SLE patients identified from the 14-year and 3-year data periods varied from 5684 to 2813 cases (99.1% to 49.0%, relative to the initial cohort). As expected, incidence increased with shorter data periods (Figure 3.3) because prevalent cases were falsely being considered as incident cases. Compared to the 2001 incidence from the 15-year data period (5.7/100,000), the estimate stayed the same when the 14-year data period was used, but increased slightly (to 6.1/100,000, 95% CI: 5.5–6.6) with the 10-year data period. With a 5-year data period, incidence increased by almost 40% (to 7.9/100,000, 95% CI: 7.3–8.5), and even more dramatically (to 17.2/100,000, 95% CI: 16.3–18.2) with the 3-year data period.

SLE prevalence estimates decreased with shorter data periods (Figure 3.4). Compared to the prevalence estimate based on the initial case definition applied within a 15-year period (65.5/100,000), prevalence decreased marginally with the 14-year data period (to 65.0/100,000, 95% CI: 63.1–66.8), and only slightly (to 61.8 cases per 100,000, 95% CI: 60.0–63.5) with the 10-year data period. However, the prevalence estimate decreased by 30% (to 46.0 cases per 100,000, 95% CI: 44.4–47.5) with the 5-year data period cohort and by 46.3% (to 35.0 cases per 100,000, 95% CI: 33.6–36.3) with the 3-year data period cohort.

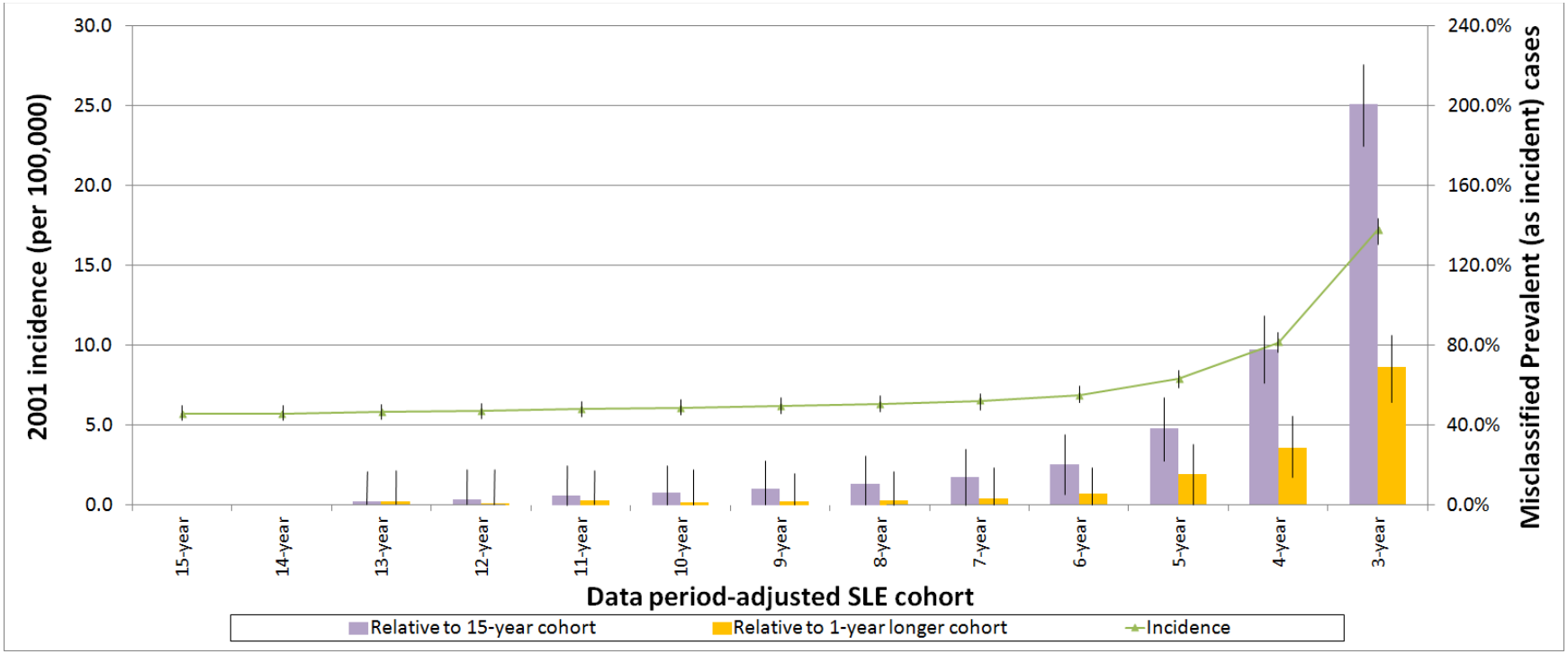

Figure 3.3: *Incidence in 2001 from data period variations.* The green line shows the 2001 incidence across data periods. The purple bar shows the percent change in misclassified prevalent (as incident) cases for a shorter data period SLE cohort relative to the 15-year data period cohort. The orange bar shows the percent change in misclassified prevalent (as incident) cases of that data period cohort relative to the cohort derived using a data period that is one year longer.

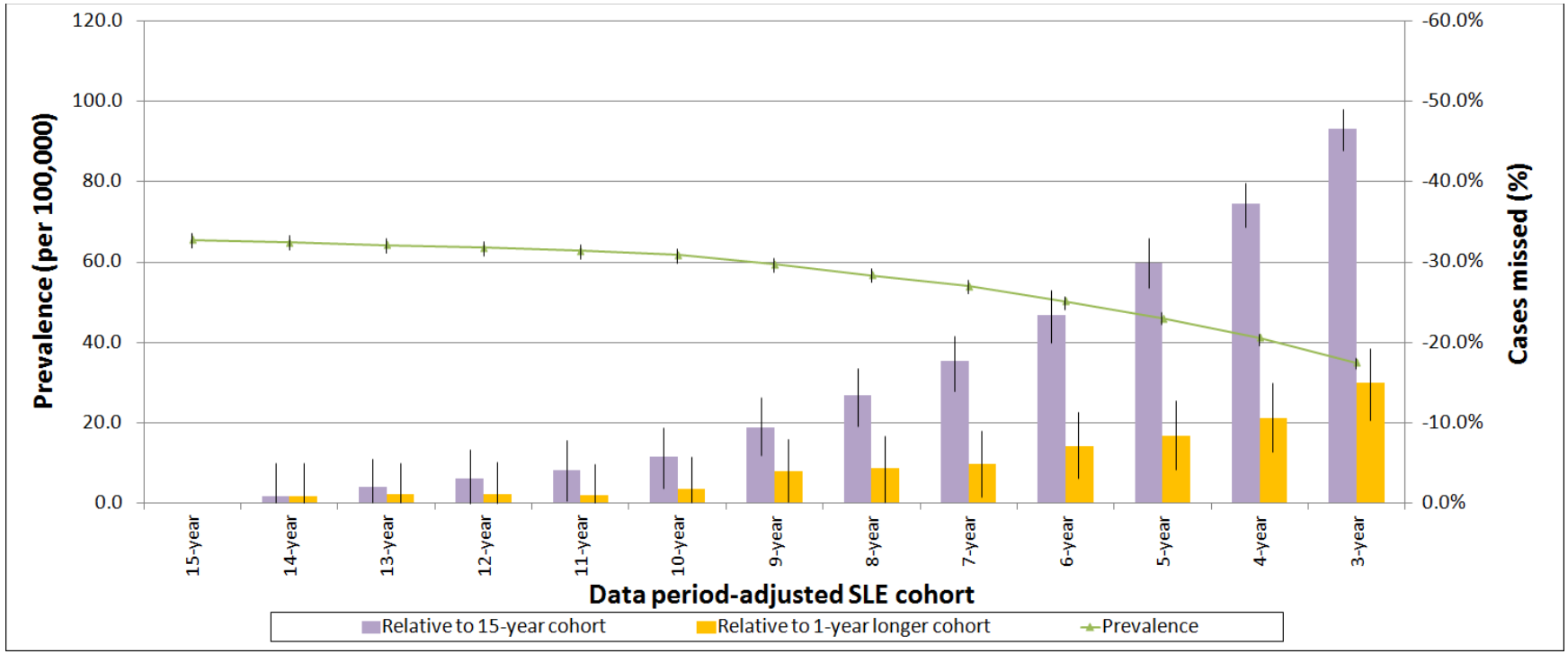

Figure 3.4: *Prevalence from data period variations*. The green line shows the prevalence across data periods. The purple bar shows the percent of cases missed for a shorter data period SLE cohort relative to the 15-year data period cohort. The orange bar shows the percent of cases missed of a particular data period cohort relative to the cohort derived using a data period that is one year longer.

Discussion

Alternative algorithms

Using variations from a 'single-rule' algorithm, incidence and prevalence estimates could vary by more than 50%. However, when all three criteria (hospitalization, rheumatology and physician definitions) were used in the algorithm, variations (1-modified multi-rule algorithms) changed the incidence and prevalence at most by about only 20%. Incidence and prevalence estimates from the algorithms requiring all three criteria, but changing two criteria at a time (2-modified multi-rule algorithms), showed more variability, as might be expected. When two alternative criteria were used, thirteen of the eighteen algorithm variants showed a statistically different incidence from the initial algorithm, and fourteen of the eighteen algorithm variants estimated a statistically different prevalence from the initial algorithm. However, the smallest and largest incidences and prevalences estimated in this study were still within the ranges of studies included in Lim *et al*.'s (2008) review[2], which is reflective of the wide variations in SLE incidence and prevalence within existing literature.

Data period variations

Incidence and prevalence estimates were affected by shorter data periods in different ways. The shorter data periods led to increases in misclassified (as incident) prevalent cases, and an inflation in the 2001 incidence estimates. On the other hand, the shorter data periods had fewer years over which to detect cases, which led to reductions in the prevalence estimates.

Estimating SLE incidence and prevalence with a 10-year data period compared to the 15-year data period did not overestimate incidence or underestimate prevalence by much. But with a 5-year data period, incidence was overestimated by almost 40% and prevalence was underestimated by almost 33% (compared to the estimates generated with 15 years of data).

Limitations

Although useful resources, one major limitation of administrative data is that they only capture data on beneficiaries obtaining medical care. Thus, individuals who may have the disease in question, but do not seek care, will not be captured. However, for a symptomatic and serious disease like SLE, this limitation may not be so important. Another limitation is that no reference standard external to the database was available. However, in many cases, rheumatic disease research and surveillance in Canada relies on administrative data without a clinical reference standard, so from this perspective, our estimates are useful because they reveal how permutations of an algorithm can increase or decrease case detection.

One limitation with the data periods examined is that some mild SLE cases may have remained undetected throughout the fifteen years, and some prevalent cases may have been misclassified as incident cases because of left censoring of the data prior to 1989. However, as the data periods became larger, there were relatively fewer misclassified and missed cases as shown by the shorter orange bars in Figures 3 and 4 from a 3-year data period to the 15-year data period. When the 14-year data period is compared to the 15-year data period, only 2% of cases were misclassified and 0.9% of cases were missed. These small values suggest that if the 15-year data period could be

compared to the 16-year data period, the number of cases misclassified or missed would be minimal.

Future directions

One future direction could be to design and evaluate new algorithms that incorporate SLE manifestation diagnosis codes like nephritis, or to use drug data from the RAMQ prescription drug claims database (though this is more complex, particularly because the database is not comprehensive) . Manifestation and drug data may improve detection within shorter data periods because the presence of these codes could indicate the presence of SLE when SLE itself is not directly captured.

Similar data is also available for other autoimmune rheumatic diseases (Sjogren's syndrome, polymyositis and scleroderma), so conducting similar analyses of algorithms with these diseases and comparing them would be of interest. Finally, using an external reference standard to validate the algorithms would be a useful, though ambitious, project.

Recommendations

This study found that incidence and prevalence estimates were dependent on the criteria used, and how the criteria were combined. Based on these results, changes to pre-existing algorithm should be made cautiously on the basis of limitations of available data and/or considerations of specific subgroups of interest. In these situations, sensitivity analyses are needed because case detection is dependent on the algorithm's criteria. In addition to the algorithm, the length of the data period available for ascertaining cases has an effect, with shorter data periods increasing incidence

estimation and decreasing prevalence estimation. Based on these results, a 10-year data period (with a washout period of at least seven years to distinguish incident from prevalent cases) appears to be acceptable for estimating incidence and prevalence for SLE. We would expect that other chronic serious relapsing and remitting diseases, like multiple sclerosis, inflammatory bowel disease, and rheumatoid arthritis would behave similarly, but additional results would be required to confirm this. Regardless of the disease being studied, careful consideration of the case ascertainment algorithm and the data period is needed for all health administrative data research.

Preamble to manuscript 2

The preceding chapter has provided some discussion of the need to carefully choose the case ascertainment algorithm and the length of time for SLE case detection. The work presented in this thesis does not feature the use of a gold standard to validate a case definition, but it is hoped that the preceding chapter does shed light on the effects of varying the case definition or the period of time under study on case selection.

In the next chapter, the initial algorithm will be used over January 1, 1996 to December 31, 2003 to ascertain SLE cases. The identified cases will form the basis of an incident cohort that will be used to study SLE manifestations, specific medications in SLE, and the potential association between SLE medication exposures and later renal manifestations.

Chapter 4: Disease characterization and medication use of a Quebec health administrative database-derived systemic lupus erythematosus (SLE) cohort

Authors: R. Ng, S. Bernatsky, E. Rahme

Abstract

Background: Systemic lupus erythematosus (SLE) is a systemic autoimmune rheumatic disease that is clinically complex to diagnose and manage. A wide array of symptoms and manifestations can appear during flares and disappear during remissions. We tracked SLE manifestations and drug use in an incident SLE cohort (number (n) = 2,010) identified from the Quebec health administrative databases across 1989–2003. We also examined the association between early antimalarial drug use and renal manifestations among cohort members covered by the RAMQ public drug insurance program.

Methods: The first-time occurrence of cutaneous, hematologic, musculoskeletal, neuropsychiatric, pulmonary, renal and serositis manifestations were identified for each patient starting from four years prior to SLE diagnosis. Time to each event was captured using Kaplan-Meier (K-M) curves. First-time use of SLE medications in a sub-cohort with RAMQ public prescription drug coverage was also examined with K-M curve analyses (n = 614; 30.5% of total cohort). The association between early antimalarial drug use and renal manifestations, was examined using multivariable Cox proportional hazards (PH) regression for a second sub-cohort of SLE patients with RAMQ coverage during SLE diagnosis (n = 791; 39.4% of total cohort).

Results: Cutaneous manifestations were the most common occurring manifestation at SLE diagnosis (30.0%, 95% confidence interval (CI): 27.7–32.2%). About two-thirds (66.2%, 95% CI: 63.4–68.9%) of patients had evidence of at least one manifestation at diagnosis, which increased to 87.2% (95% CI: 84.2–90.3%) at the end of follow-up. Regarding drug use, at the end of follow-up, 65.6% (95% CI: 56.8–74.4%) of patients used at least one SLE medication, and over half (50.3%, 95% CI: 41.8–58.8%) were prescribed prednisone at least once. After adjusting for age at diagnosis, sex, early immunosuppressant and systemic steroid drug use, other co-morbidities and how SLE was diagnosed, no statistical association was established between early use of antimalarial drugs and later renal manifestations.

Conclusion: This study provides insight regarding SLE manifestations and medication use, within a population-based sample. The vast majority of patients who were identified as having SLE according to our algorithms had other diagnostic evidence or drug use that would support an underlying diagnosis of SLE. In contrast to the dramatic protective effects that other observational studies have suggested (in small, selected clinical samples), we could not establish any protective effects for antimalarial agents against renal manifestation, in our population-based sample.

Introduction

Systemic lupus erythematosus (SLE) is an uncommon, systemic autoimmune rheumatic disease (SARD) with a wide array of symptoms and manifestations that can appear during flares and disappear during remissions[39]. Previous studies have estimated SLE incidence to be 1.0–7.6 patients per 100,000 people and prevalence to be 19.3–94.0/100,000[2]. While SLE can affect people of all ages, it is predominantly a female

disease. The female to male ratio (F:M) in clinical populations is about 9:1, but this ratio varies depending on age and race/ethnicity[139]. SLE can cause cutaneous and musculoskeletal manifestations like malar rash, discoid rash and arthritis, as well as renal, pulmonary and neuropsychiatric symptoms[39, 69]. Many of the important manifestations are summarized by the American College of Rheumatology (ACR) criteria (Table 4.1)[39, 76]. This system bases an SLE diagnosis on the cumulation of at least four of eleven criteria that can occur at any time (that is, they are not required to occur concomitantly). Many consider the date a patient meets four criteria to be the SLE diagnosis date, although others believe the clinical diagnosis of SLE predates the date when ACR criteria are fulfilled.

Canadian provincial health administrative databases are a valuable resource for health research because their wide population coverage for physician services and hospitalizations (and to some extent, publicly-insured drug prescriptions) may help in creating population-based samples that make study findings more generalizable. The comprehensiveness of administrative data also facilitates the study of uncommon diseases, like SLE. For SLE research, these databases have been used to estimate incidence and prevalence[6, 125, 131], examine associated co-morbidities[140], determine medical costs and health care use[141, 142], and study risk factors[143]. However, to our knowledge, they have not been used to examine SLE manifestations and medication use.

Criterion	Criterion Definition
1. Malar Rash	Fixed erythema, flat or raised, over the malar eminences, tending to spare the nasolabial folds
2. Discoid rash	Erythematosus raised patches with adherent keratotic scaling and follicular plugging; atrophic scarring may occur in older lesion
3. Photosensitivity	Skin rash as a result of unusual reaction to sunlight, by patient history or physician observation
4. Oral Ulcers	Oral or nasopharyngeal ulceration, usually painless, observed by physician
5. Nonerosive arthritis	Involving two or more peripheral joints, characterized by tenderness, swelling, or effusion
6. Pleuritis or pericarditis	Pleuritis - convincing history of pleuritic pain or rubbing heard by a physician or evidence of pleural effusion, OR Pericarditis - documented by electrocardiogram or rub or evidence of pericardial effusion
7. Renal disease	Persistent proteinuria > 0.5 grams per day or greater than +++ if quantification not Cellular casts - may be red cell, hemoglobin, granular, tubular or mixed
8. Neurologic disorder	Seizures - in the absence of offending drugs or known metabolic derangements; e.g. uremia, ketoacidosis, or electrolyte imbalance, OR Psychosis - in the absence of offending drugs or known metabolic derangements; e.g. uremia, ketoacidosis, or electrolyte imbalance
9. Hematologic disorder	Hemolytic anemia: with reticulocytosis, OR Leukopenia: < 4,000/mm^3 on at least 2 occasions, OR Lymphopenia: < 1,500/mm^3 on at least 2 occasions, OR Thrombocytopenia: < 100,000/mm^3 in the absence of offending drugs
10. Immunologic disorder	Anti-DNA - antibody to native DNA in abnormal titer, OR Ant-Sm - presence of antibody to Sm nuclear antigen, OR Positive finding of antiphospholipid antibodies on: i) An abnormal serum level of IgG or IgM anticardiolipid antibodies ii) A positive test result for lupus anticoagulant using a standard method, or, iii) A false-positive test result for at least 6 months confirmed by Treponema pallidum immobilization or fluorescent treponemal antibody absorption test
11. Positive Antinuclear Antibody	An abnormal titer of antinuclear antibody by immunofluorescence or an equivalent assay at any point in time and in the absence of drugs

Table 4.1: *1997 ACR classification criteria for SLE.* This system bases a SLE diagnosis on the cumulation of at least four of these criteria, at any time.

One contentious issue is the early treatment effects of antimalarials on the course of SLE. Past studies have shown a highly protective effect of antimalarials, in particular, against thrombosis (HR = 0.28, 95% CI: 0.08–0.90) and kidney (renal) damage (HR = 0.12, 95% CI: 0.02–0.97)[144, 145]. However, these highly protective effects have been challenged from several perspectives: (1) confounding by indication where antimalarial treatment was given to those with the lowest risk of kidney damage, and (2) immortal-time bias where unexposed time before exposure started is misclassified as exposed time[146-148]. In addition, these studies were mostly performed on select patients

from tertiary care centres, and may not reflect the entire spectrum of SLE patients. As well, concomitant control for other medications used in SLE was suboptimal in some of these studies. With information on SLE manifestations and medications, health administrative databases provide another means to look at the possible association between early antimalarial treatment and renal manifestations.

Objectives

There were two objectives of this study: (i) to examine longitudinally SLE manifestations and SLE medication treatment in an incident SLE cohort, and (ii) to assess the association of early antimalarial treatment and renal manifestations.

Methods

Data from January 1, 1989 to December 31, 2003 were obtained from the Maintenance et exploitation des données pour l'étude de la Clientèle hospitalière (Med-echo) hospital discharge database, the Régie de l'assurance maladie du Québec (RAMQ) physician claims database, the RAMQ prescription drug database and the RAMQ demographic file. The hospital discharge database summarizes each inpatient stay, and includes up to sixteen diagnosis fields based on International Classification of Diseases (ICD)-9 codes, including one primary discharge diagnosis for the disease most attributable to the hospitalization. The physician claims database provides information on the physician's specialization, the procedure performed and the diagnosis (ICD-9 codes). The prescription drug database contains data on when individuals were covered by public prescription drug insurance, as well prescription drugs claims data, including the dose and quantity, during public coverage. The demographic file contains personal information such as year of birth, sex and date of death. All databases are linkable by a

unique patient identifier. The demographic file, hospital discharge database, and the physician claims database cover all Quebec permanent residents while the RAMQ prescription drug database covers select groups, including all residents aged 65 years and older, those on welfare and those without private drug insurance through their employment (including self-employed individuals and those unemployed). In 2000, 25% of Quebec residents under 65 years of age were not covered by private drug insurance from their employer[149, 150].

Incident SLE cohort formation

To be included in our incident SLE cohort, an individual needed to have their first diagnosis code for SLE on a date (index date) between January 1, 1996 and December 31, 2003 while meeting at least one of the following three criteria from a previously-used algorithm[6, 125]:

1) One SLE-coded (ICD-9: 710.0) hospital discharge in the Med-echo database; and/or
2) One SLE-coded rheumatologist claim from the RAMQ database and/or
3) Two SLE-coded non-rheumatologist physician claims that are at least eight weeks apart but within two years of each other from the RAMQ database.

All selected members had at least seven years (from January 1, 1989 to December 31, 1995) without a prior SLE claim, which increased the likelihood of selecting only incident patients. Individuals were excluded if they had another SARD, either rheumatoid arthritis (ICD-9: 714.X), scleroderma (710.1), Sjogren's syndrome (710.2), polymyositis (710.4), dermatomyositis (710.3) or vasculitis (446.X and 447.6). This determination was made using a previously-evaluated algorithm that consisted of the same three criteria as in the above-stated algorithm, but with the SARD-specific ICD-9 codes replacing the SLE ICD-9 code[125, 131, 151]. The age at index date was used to classify

the patient as either pediatric onset (ages 0–18), young adult onset (ages 19–44), middle-age onset (ages 45–64) or elderly onset (ages 65 and older).

SLE manifestations

Seven groups of important SLE organ manifestations were examined: cutaneous, hematologic, musculoskeletal, neuropsychiatric, pulmonary, renal and serositis manifestations[152]. These manifestations were identified in cohort members based on the presence of at least one corresponding ICD-9 code defined for each manifestation in either the physician claims or the hospitalization discharge databases (Table 4.2). Since SLE symptoms may pre-date the actual diagnosis by several years[153], we searched from four years prior to up to eight years after the index date (SLE diagnosis). Manifestations such as photosensitivity, abnormal anti-DNA titre, presence of anti-Sm and abnormal antinuclear antibody titre cannot be identified by ICD-9 codes so were not examined. In our analyses, we only considered the first diagnostic code for each manifestation, even though physician visits and hospitalizations might continue to use that diagnostic code.

Prescription drug claim

RAMQ public drug insurance coverage periods were determined for each member by using period of coverage data. For covered individuals, their use of nine, SLE-related medications belonging to three drug types were examined. The drug types were antimalarials (chloroquine and hydroxychloroquine), immunosuppressants (azathioprine, cyclosporine, cyclophosphamide, methotrexate and mycophenolate mofetil) and systemic steroids (methylprednisolone and prednisone). A prescription claim for any of these drugs was identified through their drug identification number, a

system that uniquely identifies all drug products authorized for sale in Canada. We examined drug claims in a first SLE sub-cohort covered by public prescription drug insurance four years prior to SLE diagnosis. A second SLE sub-cohort covered by public prescription drug insurance from a period of one month prior to one month after SLE diagnosis was formed to examine the association between early antimalarial treatment and future renal manifestations.

Charlson comorbidity index

The Charlson comorbidity index is used to predict the 10-year mortality of a patient by accounting for the number and kinds of co-morbid conditions he/she has[154]. It is used in health administrative database research to adjust for a patient's over-all state of health.

Female to male (F:M) ratio

The F:M ratio is the number of female patients relative to the number of male patients, and it shows the relative number of patients between sexes.

Yearly incidence

Yearly incidence and 95% CI's were calculated from 1996–2001 by selecting all SLE patients with their index claim in a particular year and dividing their number by the population of Quebec on July 1st of the corresponding year[138, 155].

Prevalence

SLE prevalence and its 95% CI was calculated by dividing the number of all SLE cohort members alive on December 31, 2003 by Quebec's population on July 1st, 2003[138].

Statistical analyses

Estimates for the average age, F:M ratio, 1996–2001 yearly incidences and prevalence were calculated for the incident cohort. The average age and F:M ratio were also determined for the two SLE sub-cohorts.

Within the incident SLE cohort, the occurrence of each manifestation in the cohort at any time as a proportion (and 95% CI) was determined by sex and age group. The number of manifestations in the incident cohort and the number of SLE medications taken in the first SLE-sub-cohort were also determined.

Kaplan-Meier curves

Kaplan-Meier (K-M) curves were plotted for each manifestation and medication from four years prior to a maximum of eight years after the SLE index date. For K-M curves of manifestations, the whole incident cohort was eligible for inclusion, but only the first sub-cohort was used for plotting K-M curves of medication use. Individuals could also be excluded from the analysis in two other ways. For all K-M curves, individuals under 5 years of age at the index date were excluded because by virtue of their young age, follow-up could not start at four years prior to SLE diagnosis; however, only six patients were excluded from the manifestation K-M curves and two from the drug use K-M curves.

Individuals with a prior history of a particular manifestation or medication even before four years prior to SLE diagnosis were excluded from the corresponding K-M plot to minimize the likelihood that particular manifestation/medication was associated with another, prior disease condition. For manifestation K-M plots, up to 18.6% of the cohort were excluded. Similarly, for drug use K-M plots, up to 13.8% of the first SLE sub-cohort

were excluded. AN individual was censored in any K-M plot if the event did not occur by December 31, 2003 or the individual died. For all drug use K-M analyses curves, a patient was censored on the date he/she left the RAMQ prescription drug insurance plan, even if he/she rejoined later.

Cox proportional hazards (PH) regression

Cox PH regression were used to examine the association of treatment with antimalarial medication within one month of SLE diagnosis and renal manifestations among the second sub-cohort of patients (those covered by the RAMQ public drug plan at least one month prior and one month after SLE diagnosis). Individuals who used antimalarials prior to one month before SLE diagnosis were not excluded from the analysis, and their antimalarial status was dependent on their antimalarial use within one month of diagnosis. Patients who had a renal manifestation prior to or at SLE diagnosis were excluded. For this analysis, seven covariates were considered. For demographic characteristics, the patient's sex and their age group at SLE diagnosis (0-18, 19-44, 45-64, 65+) were used. A patient's concurrent use of systemic steroids and immunosuppressants within one month of SLE diagnosis was also considered. To account for co-morbidities, the individual's Charlson comorbidity index was calculated at the index date. As a proxy for SLE severity, how the SLE case was first diagnosed (either during hospitalization, by a rheumatologist, or a non-rheumatologist) was also considered. Model selection was done using Hosmer and Lemeshow's purposeful selection method[156]. Any variables that did not meet a 20% level of significance in univariable analyses were excluded from the initial multivariable model. To obtain the final multivariable model, variables included in the initial multivariable model were excluded if they were not significant at the 5% level. Interaction between sex and age

group was examined because manifestations express themselves differently depending on the sex and age of the individual[46, 67].

Manifestation	ICD-9 diagnosis code
Cutaneous manifestations	
Alopecia	704.0
Apthous Ulcer	528.2
Erythema multiforme	695.1
Erythema nodusum	695.2
Lupus erythematosus	695.4 & 695.9
Raynaud's syndrome	443.0
Hematologic manifestations	
Anemia	283.0, 284.8, 285.9
Antiphospholipid syndrome	286.5, 289.8, 289.9
Neutropenia	288.0, 288.9
Thrombocytopenia	287.3, 287.5
Musculoskeletal manifestations	
Polyarthritis	714.9, 715.9, 716.5, 716.9
Neuropsychiatric manifestations	
Epilepsy and convulsions	345.9, 780.3
Schizophrenic disorders	295.X
Episodic mood disorders	296.X
Other non-organic psychoses	298.X
Pulmonary manifestations	
Alveolar pneumopathy	516.9
Bronchiolitis Obliterans organizing pneumonia	516.8
Idiopathic fibrosins alveolitis	516.3
Pulmonary hypertension	416.0
Lung involvement from another classified disease	517.8
Renal manifestations	
Impaired renal function	588.X
Nephritis	580.X - 586.X
Proteniuria	791.0
Renal sclerosis	587.X
Serosits manifestations	
Pericarditis	420.9, 423.9
Pleuritis	511.0, 511.9

Table 4.2: *The SLE manifestations examined (with corresponding ICD-9 codes).*

Results

The complete incident SLE cohort (n = 2,010) had an F:M ratio of 4.6:1. The average age at SLE diagnosis of this cohort was 43.5 years (95% CI: 42.7–44.2) with females (42.3 years of age, 95% CI: 41.4–43.1) being diagnosed 6.7 years earlier than

men (48.9 years of age, 95% CI: 47.1–50.7). At diagnosis, 143 of the patients were pediatric (7.1%), 970 were young adults (48.3%), 612 were middle-aged (30.5%) and 285 (14.2%) were above 65 years of age. Incidence from 1996–2001 varied from 3.1–4.1/100,000 with female yearly incidence varying from 5.0–6.4/100,000 and male yearly incidence varying from 0.8–1.6/100,000. The over-all prevalence was estimated to be 24.3/100,000 (95% CI: 23.2–25.4) with female prevalence estimated at 39.9/100,000 (95% CI: 37.9–41.9) and male prevalence at 8.3/100,000 (95% CI: 7.4–9.2).

The frequency of each manifestation was tabulated by age group and sex (Table 4.3). Cutaneous manifestations were found less frequently in older age groups. A large proportion of pediatric SLE patients (as compared to older subjects) had hematologic and renal manifestations. Neuropsychiatric manifestations were more common in the elderly onset group, as opposed to the younger subjects. The occurrence of musculoskeletal manifestations was similar between sexes, across all age groups. However, across all age groups, renal manifestations were more common in males than females. For both sexes, pulmonary manifestations were the least common manifestation. Approximately 15.5% of cohort members were not diagnosed with any of the seven manifestations during follow-up (Figure 4.1).

Proportion with a manifestation (%)	Sex	Age group									
		Pediatric onset		Young adult onset		Middle-aged onset		Elderly onset		All age groups	
Cutaneous	Females	70.8	[63.1– 78.5]	58.9	[55.6– 62.3]	51.9	[47.4– 58.2]	43.4	[36.7– 50.1]	55.9	[53.5– 58.2]
	Males	76.9	[56.9– 96.9]	49.3	[41.2– 57.4]	35.7	[27.5– 47.4]	35.0	[24.8– 45.2]	42.4	[37.3– 47.4]
	Both sexes	71.3	[64.0– 78.6]	57.5	[54.4– 60.6]	48.5	[44.6– 55.6]	41.1	[35.4– 46.7]	53.4	[51.3– 55.6]
Hematologic	Females	43.8	[35.5– 52.2]	33.6	[30.4– 36.8]	36.0	[31.8– 39.9]	53.7	[46.9– 60.4]	37.6	[35.3– 39.9]
	Males	46.2	[22.5– 69.8]	33.8	[26.1– 41.5]	44.4	[35.9– 47.7]	55.0	[44.4– 65.6]	42.7	[37.6– 47.7]
	Both sexes	44.1	[36.0– 52.1]	33.6	[30.6– 36.6]	37.7	[33.9– 40.6]	54.0	[48.3– 59.8]	38.5	[36.4– 40.6]
Musculoskeletal	Females	45.4	[37.0– 53.8]	34.4	[31.2– 37.6]	47.3	[42.9– 44.1]	55.6	[48.9– 62.3]	41.7	[39.3– 44.1]
	Males	53.8	[30.2– 77.5]	38.7	[30.8– 46.6]	42.9	[34.4– 47.7]	47.5	[36.8– 58.2]	42.7	[37.6– 47.7]
	Both sexes	46.2	[38.1– 54.2]	35.1	[32.1– 38.0]	46.4	[42.5– 44.0]	53.3	[47.6– 59.1]	41.9	[39.7– 44.0]
Neuropsychiatric	Females	10.0	[4.9– 15.1]	13.8	[11.4– 16.1]	15.4	[12.2– 16.4]	20.0	[14.6– 25.4]	14.7	[13.0– 16.4]
	Males	23.1	[3.1– 43.1]	11.3	[6.1– 16.4]	13.5	[7.6– 17.4]	17.5	[9.4– 25.6]	13.9	[10.3– 17.4]
	Both sexes	11.2	[6.1– 16.3]	13.4	[11.3– 15.5]	15.0	[12.2– 16.1]	19.3	[14.8– 23.8]	14.6	[13.0– 16.1]
Pulmonary	Females	3.8	[0.6– 7.1]	1.7	[0.8– 2.6]	2.3	[0.9– 3.0]	3.4	[1.0– 5.9]	2.2	[1.5– 3.0]
	Males	0.0	[0.0– 0.0]	3.5	[0.5– 6.5]	4.0	[0.6– 6.5]	7.5	[1.9– 13.1]	4.4	[2.3– 6.5]
	Both sexes	3.5	[0.5– 6.5]	2.0	[1.1– 2.8]	2.6	[1.4– 3.3]	4.6	[2.2– 7.0]	2.6	[1.9– 3.3]
Renal	Females	30.8	[23.0– 38.6]	16.1	[13.6– 18.6]	14.4	[11.3– 20.2]	29.3	[23.1– 35.4]	18.4	[16.5– 20.2]
	Males	61.5	[38.5– 84.6]	26.1	[19.0– 33.2]	24.6	[17.2– 33.7]	36.3	[26.0– 46.5]	29.1	[24.4– 33.7]
	Both sexes	33.6	[25.9– 41.2]	17.5	[15.1– 19.9]	16.5	[13.6– 22.1]	31.2	[25.9– 36.6]	20.3	[18.5– 22.1]
Serositis	Females	23.1	[16.0– 30.2]	11.4	[9.2– 13.5]	10.7	[8.0– 14.0]	14.1	[9.4– 18.9]	12.4	[10.8– 14.0]
	Males	15.4	[0.0– 32.5]	22.5	[15.8– 29.3]	15.1	[8.9– 22.5]	17.5	[9.4– 25.6]	18.6	[14.6– 22.5]
	Both sexes	22.4	[15.7– 29.1]	13.0	[10.9– 15.1]	11.6	[9.1– 15.0]	15.1	[11.0– 19.2]	13.5	[12.0– 15.0]

Table 4.3: *The proportion of manifestations within the SLE cohort by sex and age group at any time between four years before their SLE diagnosis to eight years after diagnosis.*

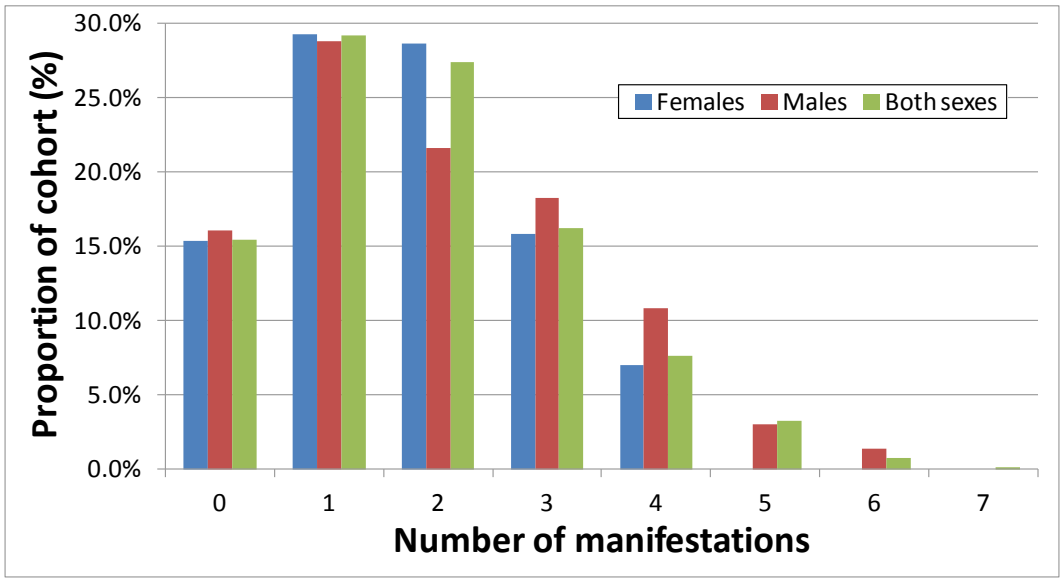

Figure 4.1: *Number of manifestations detected in SLE cohort members by sex.* The observation interval started four years before SLE diagnosis, and continued to a maximum of eight years after diagnosis.

K-M curves were plotted for each manifestation (Figure 4.2), starting at four years prior to SLE diagnosis to a maximum of eight years afterwards. In general, as the diagnosis date approached, the likelihood of accumulating a new manifestation increased, but the rate of accumulating new manifestations slowed after the SLE diagnosis date (which was based on the first ICD-9 code for SLE itself, 710.0). At SLE diagnosis, the manifestations most often diagnosed in the previous four years were cutaneous (30.0%), musculoskeletal (25.5%) and hematologic manifestations (23.9%) (Table 4.4). By the diagnosis date, 66.2% of the cohort had evidence of at least one SLE manifestation. By the end of the observation interval, 84.5% of the cohort had evidence of at least one SLE manifestation.

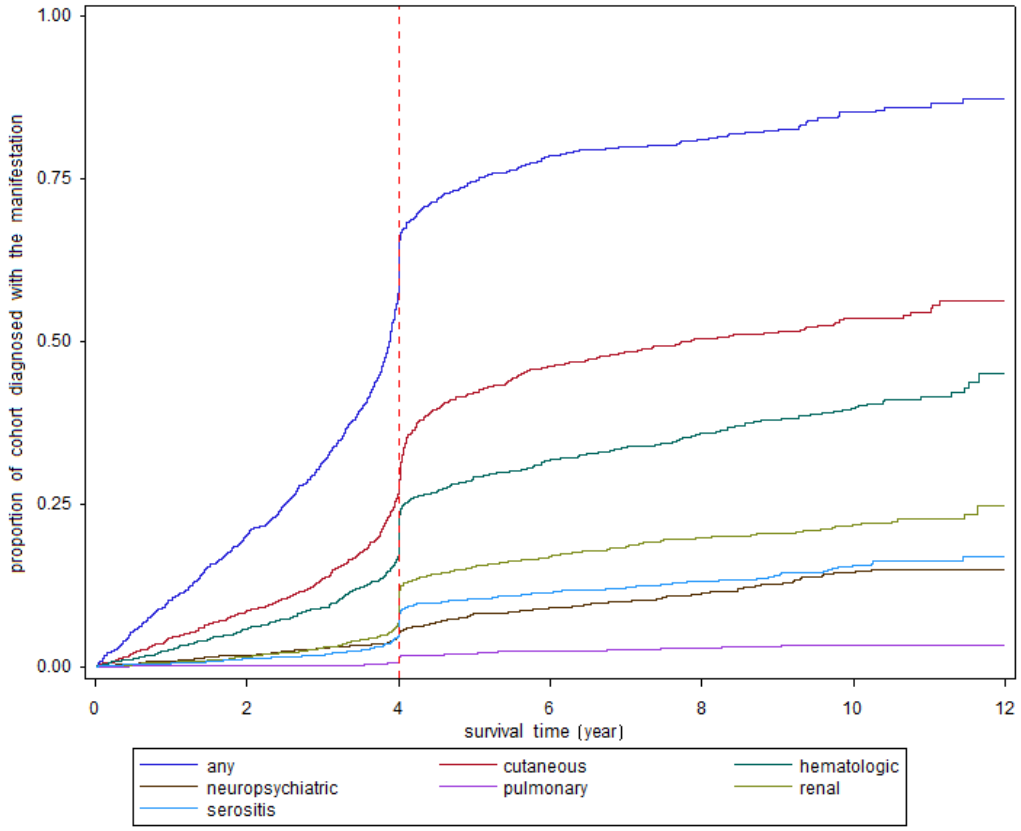

Figure 4.2: *Probability of an SLE individual not having a SLE manifestation diagnosis from four years prior to diagnosis to up to eight years after diagnosis.* The vertical dotted line denotes when SLE was diagnosed (based on the first ICD-9 code for SLE, 710.0).

Manifestation	SLE patients	Probability of being having a manifestation diagnosis starting at four years prior to diagnosis (%)				
		Two years prior	SLE diagnosis	Two years after	Five years after	Last observation
Cutaneous	1647	8.6 [7.3– 10.0]	30.0 [27.7– 32.2]	46.2 [43.7– 48.7]	51.5 [48.8– 54.2]	56.2 [52.6– 59.8]
Hematologic	1769	5.9 [4.8– 7.0]	23.9 [21.9– 25.9]	31.9 [29.6– 34.1]	38.1 [35.5– 40.6]	44.9 [40.5– 49.4]
Musculoskeletal	1637	7.8 [6.5– 9.1]	25.5 [23.4– 27.6]	34.5 [32.1– 36.9]	41.1 [38.4– 43.7]	47.3 [43.0– 51.6]
Neuropsychiatric	1877	1.8 [1.2– 2.4]	5.5 [4.5– 6.5]	9.0 [7.7– 10.4]	12.7 [10.9– 14.5]	14.9 [12.7– 17.1]
Pulmonary	2004	1.0 [0.9– 1.1]	1.5 [1.0– 2.0]	2.0 [1.4– 2.6]	3.1 [2.2– 4.0]	3.3 [2.3– 4.2]
Renal	1935	1.5 [1.0– 2.0]	12.0 [10.6– 13.5]	17.1 [15.4– 18.8]	20.7 [18.7– 22.7]	24.7 [21.0– 28.5]
Serositis	1965	1.2 [0.7– 1.7]	8.4 [7.2– 9.6]	11.3 [9.9– 12.8]	14.2 [12.4– 15.9]	16.9 [14.3– 19.6]
Any manifestation	1123	20.5 [18.1– 22.8]	66.2 [63.4– 68.9]	78.7 [76.2– 81.1]	82.6 [80.2– 85.1]	87.2 [84.2– 90.3]

Table4. 4: *Cumulative Probability of having a manifestation (and 95% CI's) at selected times relative to the SLE diagnosis date.* The column 'SLE patients' is the number of patients followed up for that particular K-M plot.

The first SLE sub-cohort (covered by the RAMQ prescription drug plan four years prior to diagnosis) consisted of 614 individuals with an F:M ratio of 4.0:1. The average age of this sub-cohort was 54.2 years (95% CI: 51.0–54.6), 10.8 years older than the complete cohort, which was not surprising given the composition of the people covered

by the public drug plan (all 65 years and over, all welfare patients and 25% of the working population). Approximately 55.1% of this sub-cohort was dispensed at least one SLE medication at any point during the twelve-year follow-up period (Figure 4.3).

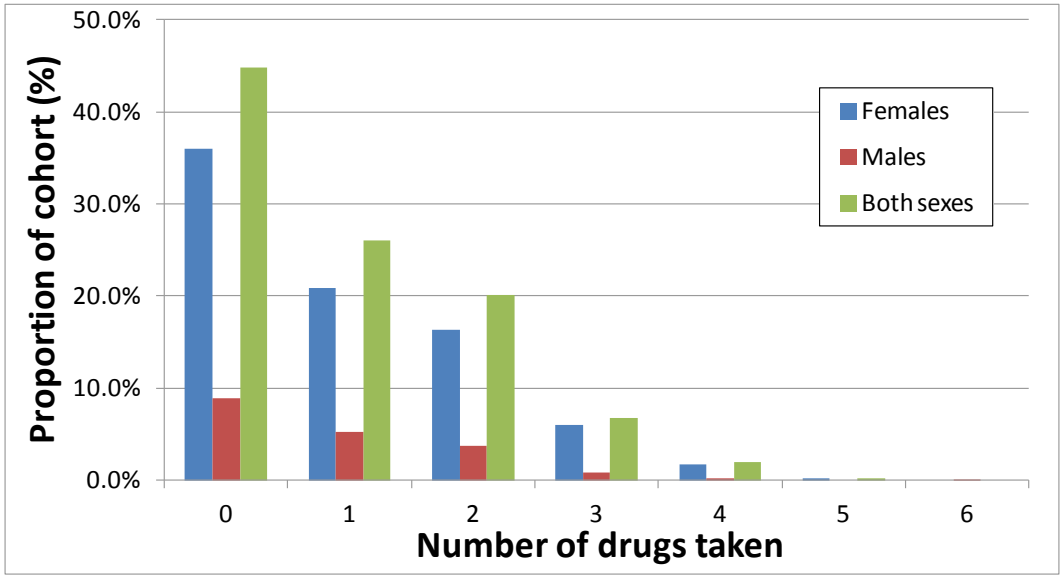

Figure 4.3: *Number of SLE cohort members with at least one recorded exposure to an SLE drug, according to sex.*

The K-M curves (Figure 4.4) show when SLE medications were first dispensed relative to SLE diagnosis for this first SLE-sub-cohort. The first-time use of these drugs of interest was most common in the period before SLE diagnosis, as shown by the increasing slope in this period. Up to the time of SLE diagnosis, the most common drug exposures of interest were prednisone (23.0%), hydroxychloroquine (18.2%) and chloroquine (2.1%) (Table 4.5). After the first few months following SLE diagnosis, first-time use gradually declined, as shown by the plateauing line.

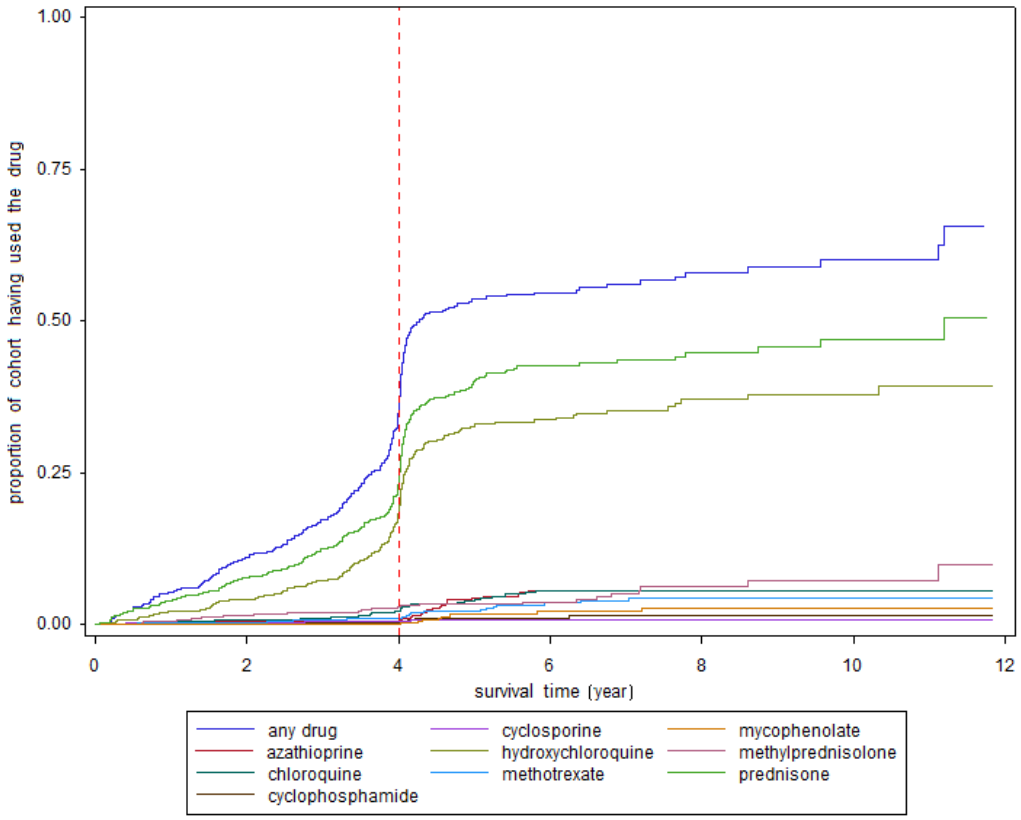

Figure 4.4: *The probability of being dispensed an SLE-related medication from four years prior to diagnosis to up to eight years after diagnosis.* The vertical dotted line denotes when SLE was diagnosed.

Manifestation	SLE patients	Probability of being prescribed an SLE medication starting at four years prior to diagnosis (%)				
		Two years prior	SLE diagnosis	Two years after	Five years after	Last observation
Azathioprine	609	0.3 [0.0– 0.8]	0.3 [0.0– 0.8]	5.4 [3.2– 7.6]	5.4 [3.2– 7.6]	5.4 [3.2– 7.6]
Chloroquine	598	0.7 [0.0– 1.4]	2.1 [0.9– 3.3]	5.1 [3.0– 7.2]	5.4 [3.3– 7.6]	5.4 [3.2– 7.6]
Cyclosporine	614	0.2 [0.0– 0.5]	0.3 [0.0– 0.8]	0.6 [0.0– 1.2]	0.6 [0.0– 1.2]	0.6 [0.0– 1.2]
Cyclophosphamide	612	0.0 [0.0– 0.0]	0.2 [0.0– 0.6]	0.8 [0.0– 1.6]	0.8 [0.0– 1.6]	1.2 [0.1– 2.3]
Hydroxychlroquine	571	4.0 [2.4– 5.7]	18.2 [14.9– 21.6]	33.5 [29.3– 37.8]	37.7 [32.7– 42.6]	39.1 [33.6– 44.7]
Methotrexate	614	0.2 [0.0– 0.5]	0.9 [0.1– 1.7]	3.4 [1.7– 5.1]	4.3 [2.2– 6.4]	4.3 [2.2– 6.4]
Methylprednisolone	596	1.4 [0.5– 2.4]	2.8 [1.4– 4.2]	4.9 [2.7– 7.2]	7.0 [3.8– 10.1]	9.8 [3.5– 16.0]
Mycophenolate mofetil	614	0.0 [0.0– 0.0]	0.0 [0.0– 0.0]	1.9 [0.6– 3.2]	5.4 [3.8– 7.1]	5.4 [5.4– 7.1]
Prednisone	529	7.5 [5.2– 9.8]	23.0 [19.2– 26.8]	42.4 [37.8– 47.1]	45.5 [40.3– 50.7]	50.3 [41.8– 58.8]
Any drug	486	10.9 [8.0– 13.7]	34.9 [30.5– 39.4]	54.5 [49.7– 59.3]	58.8 [53.4– 64.1]	65.6 [56.8– 74.4]

Table 4.5: *Cumulative probability of being dispensed SLE medications (and 95% CI's) relative to four years prior to SLE diagnosis*. The column 'SLE patients' denotes the number of patients used for that particular K-M analysis.

The second SLE sub-cohort (used to look at the association between antimalarial exposure at diagnosis and later renal manifestations) consisted of 791 patients (39.4% of the total cohort) with an F:M ratio of 4.4:1. The sub-cohort was also older than the complete cohort with the average patient being about 50.3 years of age (95% CI: 49.0,

51.7) at diagnosis. Females had an average age of 48.9 years (95% CI: 47.5–50.5) and males had an average age of 56.1 years (95% CI: 53.1–59.2). In the univariable Cox PH regression analyses, middle-aged onset SLE was not significant at the 20% level, but remained in the initial multivariable model because of a possible sex-age interaction (Figure 4.5). In the final multivariable model, there was no clear association between antimalarial treatment within one month of diagnosis and renal manifestations (HR = 1.39, 95% CI: 0.90–2.16). However, four variables were found to have an association: pediatric SLE (HR = 2.55, 95% CI: 1.35–4.83), elderly onset SLE (HR = 2.54, 95% CI: 1.70–3.80), systemic steroid dispensation at diagnosis (HR = 1.93, 95% CI: 1.27–2.94) and a hospital SLE diagnosis (HR = 3.04, 95% CI: 2.07–4.48). We were unable to detect significant interactions (at the 5% level) between sex and age group.

Variable	Univariate models	Initial multivariate model	Final multivariate model
	HR (95% CI)	HR (95% CI)	HR (95% CI)
Pediatric SLE (ages 0-18)	3.21 [1.62– 6.36]	2.27 [1.11– 4.64]	2.55 [1.35– 4.83]
Young adult onset (ages 19-44)	1.00 --	1.00 --	1.00 --
Middle-age onset (ages 45-64)	1.03 [0.59– 1.79]	0.91 [0.52– 1.60]	excluded
Elderly onset (ages 65+)	2.33 [1.45– 3.74]	2.20 [1.32– 3.65]	2.54 [1.70– 3.80]
Male	1.54 [1.01– 2.36]	1.46 [0.95– 2.26]	excluded
Female	1.00 --	1.00 --	excluded
Immunosuppressant use	3.73 [1.64– 8.50]	2.06 [0.85– 4.96]	excluded
Antimalarial use	1.64 [1.10– 2.45]	1.41 [0.91– 2.20]	1.39 [0.90– 2.16]
Systemic steroid use	2.65 [1.81– 3.88]	1.82 [1.18– 2.82]	1.93 [1.27– 2.94]
Charlson comorbidity index increase	1.09 [1.01– 1.17]	1.03 [0.96– 1.12]	excluded
Hospital diagnosis	3.60 [2.48– 5.25]	3.01 [2.03– 4.46]	3.04 [2.07– 4.48]
Rheumatologist diagnosis	0.71 [0.38– 1.35]	excluded	excluded

F**igure 4.5**: *Univariable and multivariable Cox proportional hazard regression results.*

Discussion

Men were diagnosed with SLE approximately 6.7 years later in life than women. A similar result has been seen by Nossent *et al.* (2001), who found women were

diagnosed 8.1 years earlier than men from an SLE cohort obtained using diagnosis codes from county registries and mortality databases[19]. Somers *et al.* (2007) also observed a 5.9 year diagnosis gap between females and males in an SL cohort from the UK General Practice Research Database[21]. Two possible explanations for the difference in age are that different biologic mechanisms occur between the sexes due to different hormone levels and/or the absence/presence of an extra X chromosome[16], or that females are diagnosed quicker because SLE is not suspected as early in men. Theoretically, it may also be that since women on average have more physician encounters early in adulthood than men (i.e. due to reproductive issues), they may have a greater chance of having SLE symptoms noticed and diagnosed.

Comparing the manifestation proportions of this cohort to the 1000 Canadian Faces of Lupus cohort assembled from 2005-2007[70], cutaneous (53.4% vs. 43-65%) and neuropsychiatric (14.7% vs. 8-12%) prevalences were similar, but hematologic (38.5% vs. 61-81%), musculoskeletal (41.9% vs. 64-85%), renal (20.3% vs. 40-58%) and serositis (13.5% vs. 23-45%) manifestations were somewhat lower in our cohort. This may be because we studied an 'incident' cohort while the 1000 Canadian Faces of Lupus cohort contains incident and prevalent members (thus that cohort may have had more time to accumulate different manifestations)[69].

Our study found that hematologic, renal and serositis manifestations were more prevalent in men than women (by 5.1%, 10.7% and 6.2% respectively), which is consistent with other studies of clinical cohorts[46]. We also found pediatric patients were more likely to have hematological (44.1% vs. 33.6%) and renal manifestations (33.6% vs. 17.5%) than young adult onset patients, which was also observed in the Euro-Lupus

cohort[67]. In our study, we found the middle-age onset group was more likely to have musculoskeletal manifestations, as compared to the other adult age groups. This corresponds generally to what others have published[67, 72].

K-M curves provided a graphical way to show the cumulative occurrence of SLE manifestations. These illustrate, as expected, evidence of a clustering of new-onset manifestations at the time when SLE is first diagnosed (according to our ICD-9 code algorithm for SLE). This is consistent with the clinical use of multiple ACR criteria (where at least 4 of 11 are generally required) to confirm a diagnosis of SLE. As mentioned in the introduction, some authorities feel a clinical diagnosis of SLE can predate the date a patient fulfills four ACR criteria.

One limitation of our study is the first SLE sub-group examined for drug use had an average age of onset of 50.9 years, and SLE activity and severity profiles appear to be different for middle-aged and elderly onset SLE than for young-adult SLE. Thus, it is possible that we would have seen different effects of antimalarial drugs on renal manifestations had we examined these effects only in younger patients.

To our knowledge, only one previous study has validated renal manifestations using administrative data[121]. For the other six manifestations, no validation studies have been done. While some of our manifestation prevalence estimates were similar to past work, other estimates were somewhat different (e.g. for musculoskeletal manifestations). This may be due to differences in the SLE duration and/or follow-up of patients, but a validation study of all manifestations using external chart review could quantify the sensitivity and specificity of the ICD-9 manifestation codes used. These codes could then be used as part of an ascertainment algorithm to select SLE patients,

and to identify SLE subgroups that may be at greater risk for co-morbidities, complications, resulting in potentially more health care use, and/or higher mortality.

The multivariable Cox PH analysis showed no clear association between early antimalarial treatment and future renal manifestations. This is in contrast to previous work, which has suggested dramatic protective effects for antimalarial use and subsequent renal damage[145]. The difference in the findings may be attributable to how our study classified antimalarial exposure and attempted to adjust for confounding by SLE severity. Antimalarial exposure was only considered at baseline (the time of SLE diagnosis). Exposure misclassification is possible because patients may not actually have taken the antimalarials dispensed to them. To account for confounding, the clinical setting of SLE diagnosis was used as a proxy variable for SLE severity. Despite this effort to control for SLE severity, as well as controlling for concomitant drug use, there could have been residual confounding (for example, antimalarial use may be correlated with other clinical factors that themselves predict future renal manifestations). We also did not control for smoking, which could attenuate a real effect of antimalarials on SLE activity if it was more frequent among users of antimalarial[157], or consider adherence (limitations also present in other previous studies). As well, our HR estimate for the association between early antimalarial treatment and renal disease is imprecise, so it still includes the possibility of a protective effect. Though the number of patients in these analyses is actually fairly high compared to other studies, we obviously still had inadequate power to demonstrate a clear association between antimalarial use and renal disease.

The association of early systemic steroid treatment and renal manifestations (HR = 1.93, 95% CI: 1.27–2.94) observed may be due to SLE patients with the highest risk for the most severe SLE (based on their initial symptoms) are often more likely to be prescribed prednisone. This same group who have severe non-renal (e.g. neuropsychiatric or hematologic) manifestations initially, are most likely, regardless of treatment, to later develop other cumulative manifestations (including renal manifestations)[79]. In our analysis, adjustment of initial SLE severity was attempted, but as mentioned before, there may be residual confounding.

Conclusions

The determination of some SLE manifestations from health administrative databases using diagnosis codes may be viable, but manifestation code validation studies would help further determine the validity of this approach. In contrast to dramatic potential protective effects that other observational studies (in small, selected clinical samples) have suggested, we could not establish any protective effects for antimalarial agents against renal manifestations in our population-based sample.

Chapter 5: Discussion

Chapter 3 highlighted some methodological issues of SLE case ascertainment, and chapter 5 provided an example of how administrative data may be used to examine SLE manifestations and drug use. While these two chapters considered each aspect individually, the first part of Chapter 6 will show how the SLE diagnostic codes used for case ascertainment algorithms might be complemented by the use of SLE manifestation diagnostic codes. The second part of chapter 6 will touch on limitations present in both studies. The final part of this chapter will discuss possible future SLE research with health administrative data.

SLE case ascertainment algorithms and SLE manifestations

SLE case ascertainment algorithms SLE manifestations can be used together by the researcher to better understand the data and potentially improve study results. One way they can be used together is to use SLE manifestations as an indicator of the type of SLE patients selected by the algorithm. As demonstrated in chapter 3, the case ascertainment algorithm acts as the study's case definition and determines the number of SLE patients selected. However, the stringency of the algorithm could also affect the type of SLE patient detected. An algorithm with stricter criteria that requires more SLE visits in a shorter period may select patients with more severe SLE (who require more physician care). Although SLE severity cannot be directly measured with these databases, examining the number of manifestations within an SLE cohort should be an indicator of severity where more manifestations are associated with greater SLE severity. To see if the stringency of the algorithm actually impacts the type of SLE patient selected, one can compare the frequency (across cohorts assembled based on

different case definitions) of seven manifestations (cutaneous, hematological, musculoskeletal, neuropsychiatric, pulmonary, renal and serositis) examined in chapter 5. Table 5.1 shows these manifestations tabulated for the initial SLE cohort derived (as defined in chapter 3) over a 15-year period from the algorithm of: one SLE hospitalization, one rheumatologist SLE visit, and/or two non-rheumatologist SLE visits at least 8 weeks apart but within two years ('Cohort A'). For comparison, SLE manifestations were also tabulated for an SLE cohort derived from the broadest SLE case definition algorithm (one SLE hospitalization, one SLE visit to any of the defined specialists, and/or two SLE non-defined specialist visits at any time), labeled 'Cohort B', and from the most restrictive SLE case definition (one hospitalization, three rheumatologist SLE visits and/or four non-rheumatologist SLE visits), labeled 'Cohort C'.

Manifestation	Proportion of cohort with manifestations and 95% CI's (%)		
	Cohort A	Cohort B	Cohort C
Cutaneous	67.0 [65.8 - 68.3]	61.0 [59.8 - 62.1]	75.0 [73.7 - 76.3]
Hematologic	50.4 [49.1 - 51.7]	47.1 [45.9 - 48.2]	55.7 [54.2 - 57.2]
Musculoskeletal	63.7 [62.5 - 65.0]	61.5 [60.3 - 62.6]	64.4 [62.9 - 65.8]
Neuropsychiatic	20.4 [19.3 - 21.4]	19.6 [18.6 - 20.5]	22.0 [20.7 - 23.2]
Pulmonary	6.9 [6.2 - 7.5]	6.1 [5.5 - 6.7]	7.9 [7.1 - 8.7]
Renal	32.2 [31.0 - 33.4]	29.0 [27.9 - 30.1]	38.8 [37.3 - 40.2]
Serositis	22.0 [20.9 - 23.0]	20.0 [19.0 - 20.9]	26.2 [24.9 - 27.5]

Table 5.1: *The proportion of SLE manifestations in SLE cohorts derived using three different case ascertainment algorithms.*

Table 5.1 illustrates a tendency for 'Cohort A' to have a higher proportion of these manifestations compared to 'Cohort B'. This finding supports the hypothesis that the broader, less stringent definition ('Cohort B') selects patients with less severe SLE (or possibly, subjects who do not have SLE at all). Conversely, 'Cohort A' has a lower proportion of these manifestations compared to 'Cohort C'. This finding also supports the hypothesis that a more restrictive, less stringent definition ('Cohort C') selects

patients with more severe SLE, and excludes patients with less severe SLE (and possibly those that do not have SLE).

What potential ramifications might there be if a less stringent definition is used in future studies? As an example, if 'Cohort B' was assembled using the broader, less stringent definition for a study of health care resources used by SLE patients, the average costs (per patient) may be artificially low because people without SLE might be included. Conversely, if 'Cohort C' was assembled using a more restrictive, more stringent definition for the same study, the average costs (per patient) may be artificially high because the full spectrum of SLE cases have not been included.

Limitations pertaining to both studies

In chapters 3 and 5, several limitations related to health administrative data were discussed. In addition to these limitations, others can be raised. Firstly, the physician claims database does contain missing diagnostic code data indicated by either a blank field, 'V999' or '0000'. Together, these codes made up 23.3% of all physician claims. For SLE case ascertainment, missing diagnosis codes would result in SLE cases being missed or ascertained at a later date after their actual diagnosis date. Similarly, missing diagnosis codes may have resulted in the non-detection of manifestations, or manifestations being observed at a later time. A second limitation is that if an individual who already has SLE immigrates to Quebec and is subsequently detected by the algorithm as a case, SLE incidence in Quebec will be an overestimate.

Future directions

There are many future directions possible, such as examining the occurrence of SLE co-morbidities like cardiovascular disease, osteoporosis, infections and malignancies with K-M curves, and possible associated risk factors with Cox PH regression.

When the incident cohort in chapter 5 was constructed, individuals with other systemic autoimmune rheumatic diseases (SARDs) were excluded (on the basis of physician billing and hospitalization data). At early onset, presentation of different SARDs (such as scleroderma) can be similar to SLE, and sometimes, only when the disease progresses and the symptoms become more distinct, does the final diagnosis emerge. To assess how frequently patients are given an initial diagnosis (e.g. SLE) and then later another diagnosis (e.g. scleroderma), administrative data may be quite useful. However, this topic is tricky because some individuals can have overlap syndrome where he/she actually fulfills criteria for more than one disease.

Health care use by SLE patients could also be examined in the years leading up to and after SLE diagnosis. SLE has a relapsing-remitting nature, and health care use could be analyzed to determine differences in associated costs during different periods.

Chapter 7: Conclusion

This thesis studied two SLE health administrative database research components. The first component was a methodological study of SLE through comparisons of different case ascertainment algorithms and data period lengths. When one parameter of the initial algorithm was changed, SLE incidence in 1998 varied between 4.8–6.7/100,000 from 6.0/100,000 (95% CI: 5.5–6.5) and prevalence varied between 52.1–72.3/100,000 from 65.5/100,000 (95% CI: 63.7–67.4). The case ascertainment algorithms provided even more variable incidence and prevalence estimates when two parameters were changed.

Shrinking the data period used to detect cases from fifteen to five years resulted in SLE prevalence being underestimated by 30% (46.0/100,000; 95% CI: 44.4–47.5/100,000) and the 2001 SLE incidence being overestimated by almost 40% from 5.7/100,000 to 7.9/100,000 (95% CI: 7.3–8.5/100,000). Based on these results, careful consideration of the case definition used is needed. Regarding data period length, a 10-year time window was sufficient to keep incidence and prevalence estimates to within 90% of the original estimates.

The second component of this thesis looked at SLE manifestations and medication use in an incident SLE cohort. For incident SLE patents ascertained between January 1, 1996 and December 31, 2003, the first-time occurrence of seven manifestations (cutaneous, hematological, musculoskeletal, neuropsychiatric, pulmonary, renal and serositis manifestations) was observed from four years prior to up to eight years after SLE diagnosis. At the end of follow-up, 87.2% of the cohort (n =

2010) was diagnosed with at least one manifestation, the most common manifestation being cutaneous manifestations (56.2%, 95% CI: 52.6–59.8%).

SLE medication use was described in a smaller SLE sub-cohort of those covered by the provincial prescription drug plan (n = 614). At the end of follow-up, 65.6% (95% CI: 56.8–74.4%) of individuals used at least one of the medications of interest. Of the medications examined, prednisone was most frequently claimed (50.3%, 95% CI: 41.8–58.8%). A multivariable Cox PH regression found no association between early antimalarial use and renal manifestations (HR = 1.39, 95% CI: 0.90–2.16).

In summary, in health administrative database studies, SLE case ascertainment algorithm selection is important, and the use of additional data like SLE manifestations and medication use might provide additional information for longitudinal studies.

Works Cited

1. Public Health Agency of Canada, Arthritis Consumer Experts, Arthritis Community Research & Evaluation Unit, et al. Life with Arthritis in Canada : A personal and public health challenge. 2010.

2. Lim SS, Drenkard C. Epidemiology of systemic lupus erythematosus: capturing the butterfly. *Curr Rheumatol Rep* 2008; **10**(4): 265-72.

3. Danchenko N, Satia JA, Anthony MS. Epidemiology of systemic lupus erythematosus: a comparison of worldwide disease burden. *Lupus* 2006; **15**(5): 308-18.

4. Pons-Estel GJ, Alarcon GS, Scofield L, Reinlib L, Cooper GS. Understanding the epidemiology and progression of systemic lupus erythematosus. *Semin Arthritis Rheum* 2010; **39**(4): 257-68.

5. Peschken CA, Esdaile JM. Systemic lupus erythematosus in North American Indians: a population based study. *The Journal of rheumatology* 2000; **27**(8): 1884-91.

6. Bernatsky S, Joseph L, Pineau CA, Tamblyn R, Feldman DE, Clarke AE. A population-based assessment of systemic lupus erythematosus incidence and prevalence--results and implications of using administrative data for epidemiological studies. *Rheumatology* 2007; **46**(12): 1814-8.

7. Siegel M, Holley HL, Lee SL. Epidemiologic studies on systemic lupus erythematosus. Comparative data for New York City and Jefferson County, Alabama, 1956-1965. *Arthritis and rheumatism* 1970; **13**(6): 802-11.

8. Amor B, Bouchet H, Delrieu F. [National survey on reactive arthritis by the French Society of Rheumatology]. *Rev Rhum Mal Osteoartic* 1983; **50**(11): 733-43.

9. Chakravarty EF, Bush TM, Manzi S, Clarke AE, Ward MM. Prevalence of adult systemic lupus erythematosus in California and Pennsylvania in 2000: estimates obtained using hospitalization data. *Arthritis Rheum* 2007; **56**(6): 2092-4.

10. Michet CJ, Jr., McKenna CH, Elveback LR, Kaslow RA, Kurland LT. Epidemiology of systemic lupus erythematosus and other connective tissue diseases in Rochester, Minnesota, 1950 through 1979. *Mayo Clin Proc* 1985; **60**(2): 105-13.

11. Vilar MJ, Sato EI. Estimating the incidence of systemic lupus erythematosus in a tropical region (Natal, Brazil). *Lupus* 2002; **11**(8): 528-32.

12. Iseki K, Miyasato F, Oura T, Uehara H, Nishime K, Fukiyama K. An epidemiologic analysis of end-stage lupus nephritis. *Am J Kidney Dis* 1994; **23**(4): 547-54.

13. Naleway AL, Davis ME, Greenlee RT, Wilson DA, McCarty DJ. Epidemiology of systemic lupus erythematosus in rural Wisconsin. *Lupus* 2005; **14**(10): 862-6.

14. Ward MM. Prevalence of physician-diagnosed systemic lupus erythematosus in the United States: results from the third national health and nutrition examination survey. *J Womens Health (Larchmt)* 2004; **13**(6): 713-8.

15. Hochberg MC. Prevalence of systemic lupus erythematosus in England and Wales, 1981-2. *Ann Rheum Dis* 1987; **46**(9): 664-6.

16. Lockshin MD. Sex differences in autoimmune disease. *Lupus* 2006; **15**(11): 753-6.

17. Alamanos Y, Voulgari PV, Siozos C, et al. Epidemiology of systemic lupus erythematosus in northwest Greece 1982-2001. *The Journal of rheumatology* 2003; **30**(4): 731-5.

18. Deligny C, Thomas L, Dubreuil F, et al. [Systemic lupus erythematosus in Martinique: an epidemiologic study]. *Rev Med Interne* 2002; **23**(1): 21-9.

19. Nossent HC. Systemic lupus erythematosus in the Arctic region of Norway. *The Journal of rheumatology* 2001; **28**(3): 539-46.

20. Lopez P, Mozo L, Gutierrez C, Suarez A. Epidemiology of systemic lupus erythematosus in a northern Spanish population: gender and age influence on immunological features. *Lupus* 2003; **12**(11): 860-5.

21. Somers EC, Thomas SL, Smeeth L, Schoonen WM, Hall AJ. Incidence of systemic lupus erythematosus in the United Kingdom, 1990-1999. *Arthritis and rheumatism* 2007; **57**(4): 612-8.

22. Hochberg MC. The incidence of systemic lupus erythematosus in Baltimore, Maryland, 1970-1977. *Arthritis and rheumatism* 1985; **28**(1): 80-6.

23. Hopkinson ND, Doherty M, Powell RJ. Clinical features and race-specific incidence/prevalence rates of systemic lupus erythematosus in a geographically complete cohort of patients. *Ann Rheum Dis* 1994; **53**(10): 675-80.

24. Nived O, Sturfelt G, Wollheim F. Systemic lupus erythematosus in an adult population in southern Sweden: incidence, prevalence and validity of ARA revised classification criteria. *Br J Rheumatol* 1985; **24**(2): 147-54.

25. Segasothy M, Phillips PA. Systemic lupus erythematosus in Aborigines and Caucasians in central Australia: a comparative study. *Lupus* 2001; **10**(6): 439-44.

26. Demas KL, Costenbader KH. Disparities in lupus care and outcomes. *Curr Opin Rheumatol* 2009; **21**(2): 102-9.

27. Jacobsen S, Petersen J, Ullman S, et al. A multicentre study of 513 Danish patients with systemic lupus erythematosus. I. Disease manifestations and analyses of clinical subsets. *Clin Rheumatol* 1999; **17**(6): 468-77.

28. Costallat LT, Coimbra AM. Systemic lupus erythematosus: clinical and laboratory aspects related to age at disease onset. *Clin Exp Rheumatol* 1994; **12**(6): 603-7.

29. Boddaert J, Huong DL, Amoura Z, Wechsler B, Godeau P, Piette JC. Late-onset systemic lupus erythematosus: a personal series of 47 patients and pooled analysis of 714 cases in the literature. *Medicine (Baltimore)* 2004; **83**(6): 348-59.

30. McCarty DJ, Manzi S, Medsger TA, Jr., Ramsey-Goldman R, LaPorte RE, Kwoh CK. Incidence of systemic lupus erythematosus. Race and gender differences. *Arthritis and rheumatism* 1995; **38**(9): 1260-70.

31. Johnson AE, Gordon C, Palmer RG, Bacon PA. The prevalence and incidence of systemic lupus erythematosus in Birmingham, England. Relationship to ethnicity and country of birth. *Arthritis and rheumatism* 1995; **38**(4): 551-8.

32. Bae SC, Fraser P, Liang MH. The epidemiology of systemic lupus erythematosus in populations of African ancestry: a critical review of the "prevalence gradient hypothesis". *Arthritis Rheum* 1998; **41**(12): 2091-9.

33. Westlake SL, Edwards CJ. Anti-malarials and Lupus in West Africa use and lupus in Africans. *Lupus* 2009; **18**(3): 193-5.

34. Borchers AT, Naguwa SM, Shoenfeld Y, Gershwin ME. The geoepidemiology of systemic lupus erythematosus. *Autoimmun Rev* 2010; **9**(5): A277-87.

35. Johnson AE, Gordon C, Hobbs FD, Bacon PA. Undiagnosed systemic lupus erythematosus in the community. *Lancet* 1996; **347**(8998): 367-9.

36. Terris DD, Litaker DG, Koroukian SM. Health state information derived from secondary databases is affected by multiple sources of bias. *J Clin Epidemiol* 2007; **60**(7): 734-41.

37. Patel M, Clarke AM, Bruce IN, Symmons DP. The prevalence and incidence of biopsy-proven lupus nephritis in the UK: Evidence of an ethnic gradient. *Arthritis Rheum* 2006; **54**(9): 2963-9.

38. Stahl-Hallengren C, Jonsen A, Nived O, Sturfelt G. Incidence studies of systemic lupus erythematosus in Southern Sweden: increasing age, decreasing frequency of renal manifestations and good prognosis. *The Journal of rheumatology* 2000; **27**(3): 685-91.

39. Tan EM, Cohen AS, Fries JF, et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1982; **25**(11): 1271-7.

40. Smith EL, Shmerling RH. The American College of Rheumatology criteria for the classification of systemic lupus erythematosus: strengths, weaknesses, and opportunities for improvement. *Lupus* 1999; **8**(8): 586-95.

41. Petri M, Magder L. Classification criteria for systemic lupus erythematosus: a review. *Lupus* 2004; **13**(11): 829-37.

42. Guidelines for referral and management of systemic lupus erythematosus in adults. American College of Rheumatology Ad Hoc Committee on Systemic Lupus Erythematosus Guidelines. *Arthritis Rheum* 1999; **42**(9): 1785-96.

43. Hughes GR. Is it lupus? The St. Thomas' Hospital "alternative" criteria. *Clin Exp Rheumatol* 1998; **16**(3): 250-2.

44. Lim SS, Drenkard C, McCune WJ, et al. Population-based lupus registries: advancing our epidemiologic understanding. *Arthritis Rheum* 2009; **61**(10): 1462-6.

45. Uramoto KM, Michet CJ, Jr., Thumboo J, Sunku J, O'Fallon WM, Gabriel SE. Trends in the incidence and mortality of systemic lupus erythematosus, 1950-1992. *Arthritis and rheumatism* 1999; **42**(1): 46-50.

46. Lu LJ, Wallace DJ, Ishimori ML, Scofield RH, Weisman MH. Review: Male systemic lupus erythematosus: a review of sex disparities in this disease. *Lupus* 2010; **19**(2): 119-29.

47. Verthelyi D, Ansar Ahmed S. Characterization of estrogen-induced autoantibodies to cardiolipin in non-autoimmune mice. *J Autoimmun* 1997; **10**(2): 115-25.

48. Roubinian JR, Talal N, Greenspan JS, Goodman JR, Siiteri PK. Effect of castration and sex hormone treatment on survival, anti-nucleic acid antibodies, and glomerulonephritis in NZB/NZW F1 mice. *J Exp Med* 1978; **147**(6): 1568-83.

49. Lahita RG, Bradlow HL. Klinefelter's syndrome: hormone metabolism in hypogonadal males with systemic lupus erythematosus. *J Rheumatol Suppl* 1987; **14 Suppl 13**: 154-7.

50. Costenbader KH, Feskanich D, Stampfer MJ, Karlson EW. Reproductive and menopausal factors and risk of systemic lupus erythematosus in women. *Arthritis Rheum* 2007; **56**(4): 1251-62.

51. Buyon JP, Petri MA, Kim MY, et al. The effect of combined estrogen and progesterone hormone replacement therapy on disease activity in systemic lupus erythematosus: a randomized trial. *Ann Intern Med* 2005; **142**(12 Pt 1): 953-62.

52. Moorthy LN, Peterson MG, Onel KB, Lehman TJ. Do children with lupus have fewer male siblings? *Lupus* 2008; **17**(2): 128-31.

53. Grennan DM, Parfitt A, Manolios N, et al. Family and twin studies in systemic lupus erythematosus. *Dis Markers* 1997; **13**(2): 93-8.

54. Deapen D, Escalante A, Weinrib L, et al. A revised estimate of twin concordance in systemic lupus erythematosus. *Arthritis Rheum* 1992; **35**(3): 311-8.

55. Sestak AL, Shaver TS, Moser KL, Neas BR, Harley JB. Familial aggregation of lupus and autoimmunity in an unusual multiplex pedigree. *J Rheumatol* 1999; **26**(7): 1495-9.

56. Sestak AL, Nath SK, Sawalha AH, Harley JB. Current status of lupus genetics. *Arthritis Res Ther* 2007; **9**(3): 210.

57. James JA, Kaufman KM, Farris AD, Taylor-Albert E, Lehman TJ, Harley JB. An increased prevalence of Epstein-Barr virus infection in young patients suggests a possible etiology for systemic lupus erythematosus. *J Clin Invest* 1997; **100**(12): 3019-26.

58. Costenbader KH, Kim DJ, Peerzada J, et al. Cigarette smoking and the risk of systemic lupus erythematosus: a meta-analysis. *Arthritis Rheum* 2004; **50**(3): 849-57.

59. Parks CG, Cooper GS, Nylander-French LA, et al. Occupational exposure to crystalline silica and risk of systemic lupus erythematosus: a population-based, case-control study in the southeastern United States. *Arthritis Rheum* 2002; **46**(7): 1840-50.

60. McGrath H, Jr. Ultraviolet-A1 irradiation decreases clinical disease activity and autoantibodies in patients with systemic lupus erythematosus. *Clin Exp Rheumatol* 1994; **12**(2): 129-35.

61. Parks CG, Cooper GS. Occupational exposures and risk of systemic lupus erythematosus. *Autoimmunity* 2005; **38**(7): 497-506.

62. Cooper GS, Wither J, Bernatsky S, et al. Occupational and environmental exposures and risk of systemic lupus erythematosus: silica, sunlight, solvents. *Rheumatology (Oxford)* 2010; **49**(11): 2172-80.

63. Cooper GS, Parks CG, Treadwell EL, St Clair EW, Gilkeson GS, Dooley MA. Occupational risk factors for the development of systemic lupus erythematosus. *J Rheumatol* 2004; **31**(10): 1928-33.

64. Wang J, Pan HF, Ye DQ, Su H, Li XP. Moderate alcohol drinking might be protective for systemic lupus erythematosus: a systematic review and meta-analysis. *Clin Rheumatol* 2008; **27**(12): 1557-63.

65. Cooper GS, Dooley MA, Treadwell EL, St Clair EW, Gilkeson GS. Hormonal and reproductive risk factors for development of systemic lupus erythematosus: results of a population-based, case-control study. *Arthritis Rheum* 2002; **46**(7): 1830-9.

66. Barr SG, Zonana-Nacach A, Magder LS, Petri M. Patterns of disease activity in systemic lupus erythematosus. *Arthritis Rheum* 1999; **42**(12): 2682-8.

67. Cervera R, Abarca-Costalago M, Abramovicz D, et al. Systemic lupus erythematosus in Europe at the change of the millennium: lessons from the "Euro-Lupus Project". *Autoimmun Rev* 2006; **5**(3): 180-6.

68. Alarcon GS, McGwin G, Jr., Petri M, Reveille JD, Ramsey-Goldman R, Kimberly RP. Baseline characteristics of a multiethnic lupus cohort: PROFILE. *Lupus* 2002; **11**(2): 95-101.

69. Alarcon GS, McGwin G, Jr., Roseman JM, et al. Systemic lupus erythematosus in three ethnic groups. XIX. Natural history of the accrual of the American College of Rheumatology criteria prior to the occurrence of criteria diagnosis. *Arthritis Rheum* 2004; **51**(4): 609-15.

70. Peschken CA, Katz SJ, Silverman E, et al. The 1000 Canadian faces of lupus: determinants of disease outcome in a large multiethnic cohort. *J Rheumatol* 2009; **36**(6): 1200-8.

71. Macdermott EJ, Adams A, Lehman TJ. Systemic lupus erythematosus in children: current and emerging therapies. *Lupus* 2007; **16**(8): 677-83.

72. Rovensky J, Tuchynova A. Systemic lupus erythematosus in the elderly. *Autoimmun Rev* 2008; **7**(3): 235-9.

73. Cervera R, Khamashta MA, Font J, et al. Systemic lupus erythematosus: clinical and immunologic patterns of disease expression in a cohort of 1,000 patients. The European Working Party on Systemic Lupus Erythematosus. *Medicine (Baltimore)* 1993; **72**(2): 113-24.

74. Font J, Cervera R, Espinosa G, et al. Systemic lupus erythematosus (SLE) in childhood: analysis of clinical and immunological findings in 34 patients and comparison with SLE characteristics in adults. *Ann Rheum Dis* 1998; **57**(8): 456-9.

75. Sibbitt WL, Jr., Brandt JR, Johnson CR, et al. The incidence and prevalence of neuropsychiatric syndromes in pediatric onset systemic lupus erythematosus. *J Rheumatol* 2002; **29**(7): 1536-42.

76. Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1997; **40**(9): 1725.

77. Anstey NM, Bastian I, Dunckley H, Currie BJ. Systemic lupus erythematosus in Australian aborigines: high prevalence, morbidity and mortality. *Aust N Z J Med* 1993; **23**(6): 646-51.

78. Francis L, Perl A. Pharmacotherapy of systemic lupus erythematosus. *Expert Opin Pharmacother* 2009; **10**(9): 1481-94.

79. Bernatsky S, Peschken C, Fortin PR, et al. Medication use in systemic lupus erythematosus. *J Rheumatol* 2011; **38**(2): 271-4.

80. Abu-Shakra M, Urowitz MB, Gladman DD, Gough J. Mortality studies in systemic lupus erythematosus. Results from a single center. I. Causes of death. *J Rheumatol* 1995; **22**(7): 1259-64.

81. Bernatsky S, Boivin JF, Joseph L, et al. Mortality in systemic lupus erythematosus. *Arthritis Rheum* 2006; **54**(8): 2550-7.

82. Kasitanon N, Magder LS, Petri M. Predictors of survival in systemic lupus erythematosus. *Medicine (Baltimore)* 2006; **85**(3): 147-56.

83. Urowitz MB, Bookman AA, Koehler BE, Gordon DA, Smythe HA, Ogryzlo MA. The bimodal mortality pattern of systemic lupus erythematosus. *Am J Med* 1976; **60**(2): 221-5.

84. Nossent J, Cikes N, Kiss E, et al. Current causes of death in systemic lupus erythematosus in Europe, 2000--2004: relation to disease activity and damage accrual. *Lupus* 2007; **16**(5): 309-17.

85. Stoll T, Seifert B, Isenberg DA. SLICC/ACR Damage Index is valid, and renal and pulmonary organ scores are predictors of severe outcome in patients with systemic lupus erythematosus. *Br J Rheumatol* 1996; **35**(3): 248-54.

86. Rahman P, Gladman DD, Urowitz MB, Hallett D, Tam LS. Early damage as measured by the SLICC/ACR damage index is a predictor of mortality in systemic lupus erythematosus. *Lupus* 2001; **10**(2): 93-6.

87. Alarcon GS, Roseman JM, McGwin G, Jr., et al. Systemic lupus erythematosus in three ethnic groups. XX. Damage as a predictor of further damage. *Rheumatology (Oxford)* 2004; **43**(2): 202-5.

88. Sutcliffe N, Clarke AE, Gordon C, Farewell V, Isenberg DA. The association of socio-economic status, race, psychosocial factors and outcome in patients with systemic lupus erythematosus. *Rheumatology (Oxford)* 1999; **38**(11): 1130-7.

89. Abu-Shakra M, Urowitz MB, Gladman DD, Gough J. Mortality studies in systemic lupus erythematosus. Results from a single center. II. Predictor variables for mortality. *J Rheumatol* 1995; **22**(7): 1265-70.

90. Reveille JD, Bartolucci A, Alarcon GS. Prognosis in systemic lupus erythematosus. Negative impact of increasing age at onset, black race, and thrombocytopenia, as well as causes of death. *Arthritis Rheum* 1990; **33**(1): 37-48.

91. Manger K, Manger B, Repp R, et al. Definition of risk factors for death, end stage renal disease, and thromboembolic events in a monocentric cohort of 338 patients with systemic lupus erythematosus. *Ann Rheum Dis* 2002; **61**(12): 1065-70.

92. Voss A, Green A, Junker P. Systemic lupus erythematosus in Denmark: clinical and epidemiological characterization of a county-based cohort. *Scand J Rheumatol* 1998; **27**(2): 98-105.

93. Alarcon GS, Friedman AW, Straaton KV, et al. Systemic lupus erythematosus in three ethnic groups: III. A comparison of characteristics early in the natural history of the LUMINA cohort. LUpus in MInority populations: NAture vs. Nurture. *Lupus* 1999; **8**(3): 197-209.

94. Alarcon GS, Calvo-Alen J, McGwin G, Jr., et al. Systemic lupus erythematosus in a multiethnic cohort: LUMINA XXXV. Predictive factors of high disease activity over time. *Ann Rheum Dis* 2006; **65**(9): 1168-74.

95. Ward MM, Pyun E, Studenski S. Long-term survival in systemic lupus erythematosus. Patient characteristics associated with poorer outcomes. *Arthritis Rheum* 1995; **38**(2): 274-83.

96. Ward MM. Education level and mortality in systemic lupus erythematosus (SLE): evidence of underascertainment of deaths due to SLE in ethnic minorities with low education levels. *Arthritis Rheum* 2004; **51**(4): 616-24.

97. Manzi S, Meilahn EN, Rairie JE, et al. Age-specific incidence rates of myocardial infarction and angina in women with systemic lupus erythematosus: comparison with the Framingham Study. *Am J Epidemiol* 1997; **145**(5): 408-15.

98. Lee C, Ramsey-Goldman R. Bone health and systemic lupus erythematosus. *Curr Rheumatol Rep* 2005; **7**(6): 482-9.

99. Yee CS, Crabtree N, Skan J, et al. Prevalence and predictors of fragility fractures in systemic lupus erythematosus. *Ann Rheum Dis* 2005; **64**(1): 111-3.

100. Ramsey-Goldman R, Dunn JE, Huang CF, et al. Frequency of fractures in women with systemic lupus erythematosus: comparison with United States population data. *Arthritis Rheum* 1999; **42**(5): 882-90.

101. Cervera R, Khamashta MA, Font J, et al. Morbidity and mortality in systemic lupus erythematosus during a 10-year period: a comparison of early and late manifestations in a cohort of 1,000 patients. *Medicine (Baltimore)* 2003; **82**(5): 299-308.

102. Mellemkjaer L, Andersen V, Linet MS, Gridley G, Hoover R, Olsen JH. Non-Hodgkin's lymphoma and other cancers among a cohort of patients with systemic lupus erythematosus. *Arthritis Rheum* 1997; **40**(4): 761-8.

103. Tamblyn R, Lavoie G, Petrella L, Monette J. The use of prescription claims databases in pharmacoepidemiological research: the accuracy and comprehensiveness of the prescription claims database in Quebec. *J Clin Epidemiol* 1995; **48**(8): 999-1009.

104. Tamblyn R, McLeod P, Hanley JA, Girard N, Hurley J. Physician and practice characteristics associated with the early utilization of new prescription drugs. *Med Care* 2003; **41**(8): 895-908.

105. Durand M, Sheehy O, Baril JG, Lelorier J, Tremblay CL. Association between HIV infection, antiretroviral therapy and risk of acute myocardial infarction: a cohort and nested case-control study using Quebec's public health insurance database (RAMQ). *J Acquir Immune Defic Syndr* 2011.

106. Quach C, Collet JP, LeLorier J. Acute otitis media in children: a retrospective analysis of physician prescribing patterns. *Br J Clin Pharmacol* 2004; **57**(4): 500-5.

107. Lachaine J, Beauchemin C, Landry PA. Clinical and economic characteristics of patients with fibromyalgia syndrome. *Clin J Pain* 2010; **26**(4): 284-90.

108. Blais L, Boucher JM, Couture J, Rahme E, LeLorier J. Impact of a cost-sharing drug insurance plan on drug utilization among older people. *J Am Geriatr Soc* 2001; **49**(4): 410-4.

109. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol* 2005; **58**(4): 323-37.

110. Wilchesky M, Tamblyn RM, Huang A. Validation of diagnostic codes within medical services claims. *J Clin Epidemiol* 2004; **57**(2): 131-41.

111. Demers V, Melo M, Jackevicius C, et al. Comparison of provincial prescription drug plans and the impact on patients' annual drug expenditures. *CMAJ* 2008; **178**(4): 405-9.

112. Bernatsky S, Joseph L, Pineau CA, et al. Estimating the prevalence of polymyositis and dermatomyositis from administrative data: age, sex and regional differences. *Ann Rheum Dis* 2009; **68**(7): 1192-6.

113. Régie de l'assurance maladie du Québec. Liste des professionnels de la santé non-participants ou désengagés au régime de l'assurance maladie du Québec avec adresse de pratique au Québec. Quebec; 2011.

114. The College of Family Physicians of Canada, Canadiam Medical Association, The Royal College of Physicians and Surgeons of Canada. 2010 National Physician Survey (NPS): NATIONAL DEMOGRAPHICS. 2010.

115. Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care* 2005; **43**(11): 1130-9.

116. Szumski NR, Cheng EM. Optimizing algorithms to identify Parkinson's disease cases within an administrative database. *Mov Disord* 2009; **24**(1): 51-6.

117. Tricco AC, Pham B, Rawson NS. Manitoba and Saskatchewan administrative health care utilization databases are used differently to answer epidemiologic research questions. *J Clin Epidemiol* 2008; **61**(2): 192-7.

118. Rawson NS, D'Arcy C. Assessing the validity of diagnostic information in administrative health care utilization data: experience in Saskatchewan. *Pharmacoepidemiol Drug Saf* 1998; **7**(6): 389-98.

119. Bernatsky S, Joseph L, Pineau CA, Belisle P, Hudson M, Clarke AE. Scleroderma prevalence: demographic variations in a population-based sample. *Arthritis Rheum* 2009; **61**(3): 400-4.

120. Bernstein CN, Blanchard JF, Rawsthorne P, Wajda A. Epidemiology of Crohn's disease and ulcerative colitis in a central Canadian province: a population-based study. *Am J Epidemiol* 1999; **149**(10): 916-24.

121. Chibnik LB, Massarotti EM, Costenbader KH. Identification and validation of lupus nephritis cases using administrative data. *Lupus* 2010; **19**(6): 741-3.

122. Eide MJ, Krajenta R, Johnson D, et al. Identification of patients with nonmelanoma skin cancer using health maintenance organization claims data. *Am J Epidemiol* 2010; **171**(1): 123-8.

123. Hux JE, Ivis F, Flintoft V, Bica A. Diabetes in Ontario: determination of prevalence and incidence using a validated administrative data algorithm. *Diabetes Care* 2002; **25**(3): 512-6.

124. Marrie RA, Yu N, Blanchard J, Leung S, Elliott L. The rising prevalence and changing age distribution of multiple sclerosis in Manitoba. *Neurology* 2010; **74**(6): 465-71.

125. Bernatsky S, Linehan T, Hanly JG. The accuracy of administrative data diagnoses of systemic autoimmune rheumatic diseases. *J Rheumatol* 2011.

126. Lix LM, Yogendran MS, Leslie WD, et al. Using multiple data features improved the validity of osteoporosis case ascertainment from administrative databases. *J Clin Epidemiol* 2008; **61**(12): 1250-60.

127. Robinson JR, Young TK, Roos LL, Gelskey DE. Estimating the burden of disease. Comparing administrative data and self-reports. *Med Care* 1997; **35**(9): 932-47.

128. Singh JA. Discordance between self-report of physician diagnosis and administrative database diagnosis of arthritis and its predictors. *J Rheumatol* 2009; **36**(9): 2000-8.

129. Guttmann A, Nakhla M, Henderson M, et al. Validation of a health administrative data algorithm for assessing the epidemiology of diabetes in Canadian children. *Pediatr Diabetes* 2010; **11**(2): 122-8.

130. MacLean CH, Louie R, Leake B, et al. Quality of care for patients with rheumatoid arthritis. *JAMA* 2000; **284**(8): 984-92.

131. Bernatsky S, Lix L, Hanly JG, et al. Surveillance of systemic autoimmune rheumatic diseases using administrative data. *Rheumatol Int* 2011; **31**(4): 549-54.

132. Lacaille D, Guh DP, Abrahamowicz M, Anis AH, Esdaile JM. Use of nonbiologic disease-modifying antirheumatic drugs and risk of infection in patients with rheumatoid arthritis. *Arthritis Rheum* 2008; **59**(8): 1074-81.

133. Ladouceur M, Rahme E, Pineau CA, Joseph L. Robustness of prevalence estimates derived from misclassified data from administrative databases. *Biometrics* 2007; **63**(1): 272-9.

134. Katz JN, Barrett J, Liang MH, et al. Sensitivity and positive predictive value of Medicare Part B physician claims for rheumatologic diagnoses and procedures. *Arthritis Rheum* 1997; **40**(9): 1594-600.

135. Lim SS, Jamal A, Bayakly R, Tong L, Drenkard C. Lupus Registry — accuracy of hospital discharge data in identifying systemic lupus erythematosus [abstract]. *Arthritis Rheum* 2007; **54**(Suppl): S505.

136. Wijeysundera DN, Austin PC, Hux JE, Beattie WS, Buckley DN, Laupacis A. Development of an algorithm to identify preoperative medical consultations using administrative data. *Med Care* 2009; **47**(12): 1258-64.

137. Raina PS, Kirkland SA, Wolfson C, et al. Accessing health care utilization databases for health research: a Canadian longitudinal study on aging feasibility study. *Can J Aging* 2009; **28**(3): 287-94.

138. Statistics Canada. Estimates of total population for census divisions (CDs): Quebec, 2001-2004 (2004 Annual Demographic Statistics) [machine readable data file]. Ottawa, ON: Statistics Canada [producer(s)] ,Depository Services Program [distributors(s)]; 2005.

139. Barnabe C, Faris PD, Quan H. Canadian pregnancy outcomes in rheumatoid arthritis and systemic lupus erythematosus. *Int J Rheumatol* 2011; **2011**: 345727.

140. Molina MJ, Mayor AM, Franco AE, Morell CA, Lopez MA, Vila LM. Prevalence of systemic lupus erythematosus and associated comorbidities in Puerto Rico. *J Clin Rheumatol* 2007; **13**(4): 202-4.

141. Li T, Carls GS, Panopalis P, Wang S, Gibson TB, Goetzel RZ. Long-term medical costs and resource utilization in systemic lupus erythematosus and lupus nephritis: a five-year analysis of a large medicaid population. *Arthritis Rheum* 2009; **61**(6): 755-63.

142. Pelletier EM, Ogale S, Yu E, Brunetta P, Garg J. Economic outcomes in patients diagnosed with systemic lupus erythematosus with versus without nephritis: results from an analysis of data from a US claims database. *Clin Ther* 2009; **31**(11): 2653-64.

143. Schoonen WM, Thomas SL, Somers EC, et al. Do selected drugs increase the risk of lupus? A matched case-control study. *Br J Clin Pharmacol* 2010; **70**(4): 588-96.

144. Ruiz-Irastorza G, Egurbide MV, Pijoan JI, et al. Effect of antimalarials on thrombosis and survival in patients with systemic lupus erythematosus. *Lupus* 2006; **15**(9): 577-83.

145. Pons-Estel GJ, Alarcon GS, McGwin G, Jr., et al. Protective effect of hydroxychloroquine on renal damage in patients with lupus nephritis: LXV, data from a multiethnic US cohort. *Arthritis Rheum* 2009; **61**(6): 830-9.

146. Suissa S. Immortal time bias in pharmaco-epidemiology. *Am J Epidemiol* 2008; **167**(4): 492-9.

147. Vlad SC. Protective effect of hydroxychloroquine on renal damage may be biased: comment on the article by Pons-Estel et al. *Arthritis Rheum* 2009; **61**(11): 1614; author reply 5-6.

148. Vinet E, Bernatsky S, Suissa S. Have some beneficial effects of hydroxychloroquine been overestimated? Potential biases in observational studies of drug effects: comment on the article by Pons-Estel et al. *Arthritis Rheum* 2009; **61**(11): 1614-5; author reply 5-6.

149. Régie de l'assurance maladie du Québec. PA.01 - Nombre de personnes inscrites et admissibles au régime d'assurance maladie du Québec selon le sexe, le groupe d'âge et la région sociosanitaire. 2011.

150. Régie de l'assurance maladie du Québec. AM.03 - Nombre d'adhérents selon le sexe, le groupe d'âge et la région sociosanitaire de la personne assurée 2011.

151. Lacaille D, Anis AH, Guh DP, Esdaile JM. Gaps in care for rheumatoid arthritis: a population study. *Arthritis Rheum* 2005; **53**(2): 241-8.

152. O'Neill S, Cervera R. Systemic lupus erythematosus. *Best Pract Res Clin Rheumatol* 2010; **24**(6): 841-55.

153. Doria A, Zen M, Canova M, et al. SLE diagnosis and treatment: when early is early. *Autoimmun Rev* 2010; **10**(1): 55-60.

154. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *Journal of Clinical Epidemiology* 1992; **45**(6): 613-9.

155. Statistics Canada. Estimates of total population for census divisions (CDs): Quebec, 1986-2001 (2004 Annual Demographic Statistics) [machine readable data file]. Ottawa, ON: Statistics Canada [producer(s)], Depository Services Program [distributor(s)]; 2005.

156. Hosmer DW, Lemeshow S, May S. Applied Survival Analysis: Regression Modeling of Time to Event Data: John Wiley & Sons; 2011.

157. Dutz J, Werth VP. Cigarette smoking and response to antimalarials in cutaneous lupus erythematosus patients: evolution of a dogma. *J Invest Dermatol* 2011; **131**(10): 1968-70.

Appendix A: SLE incidence and prevalence studies

Study Author	Case definition	Sources for case ascertainment	Period	Country	Location	Population
1. Siegel *et al.* (1970)	Multiple systemic manifestations in the absence of other apparent causes	Hospital Files	1956-1965	USA	Jefferson County New York City	643,800 530,900
2. Fessel (1974)	≥4 1971 ARA criteria by medical chart review	Outpatient internist and dermatologist referral	1965-1973	USA	San Francisco	121,444
3. Amor *et al.* (1983)	--not in abstract--	Physician survey	1982-1982	France	Country-wide	not provided
4. Hart *et al.* (1983)	Fries and Holman criteria	Hospital medical records, outpatient physician referral and government death records	1975-1980	New Zealand	Auckland	797,367
5. Helve (1985)	All patients with an SLE diagnosis	Hospital and mental asylum computer databases	1972-1978	Finland	Finland	4,758,000
6. Nived *et al.* (1985)	Clinical diagnosis of SLE	Hospital registry and outpatient files	1981-1982	Sweden	Lund and Orup	156,924
7. Hochberg (1985)	≥4 1971 ARA criteria by medical chart review	Hospital discharge for SLE	1970-1977	USA	Baltimore	not provided
8. Michet *et al.* (1985)	≥4 1982 ACR criteria by medical chart review	Community diagnostic retrieval system for SLE, ANA, LE cell and false-positive syphilis	1950-1979	USA	Rochester	56,447
9. Hochberg (1987)	ICD-9 rubric for SLE (710.0)	Third National Morbidity Survey (1981-1982)	1981-1982	United Kingdom	England and Wales	332,270
10. Samanta *et al.* (1989)	Satisfy diagnostic criteria for SLE	Hospitals, pathology histological reports, death certificates and physician referral	1979-1988	United Kingdom	Leicaster	not provided
11. Maskarinec and Katz (1989)	--not in abstract--	Medical facilities and patient support group	1989-1989	USA	Hawaii	not provided
12. Gudmundsson and Steinsson	≥4 1982 ACR criteria	Centralized hospital registry and physician referral	1975-1984	Iceland	Iceland	239,498
13. Jonsson *et al.* (1990)	Physician diagnosis	Private and public physician referral, diagnostic registry and laboratory database	1981-1986	Sweden	Southern Sweden	200,470
14. Boyer *et al.* (1991)	≥4 1982 ACR criteria by medical record review	Alaska Area Native Health Service (AANHS) computer database	1970-1984	USA	Alaska	not provided

Study Author	Case definition	Sources for case ascertainment	Period	Country	Location	Population
15. Nossent (1992)	≥4 1982 ACR criteria	Patient records	1980-1989	Curacao	Country-wide	146,500
16. Anstey *et al.* (1993)	≥4 1982 ACR criteria	Laboratory records and medical records from regional hospitals, medical units and specialists	1984-1991	Australia	Darwin, Katherine & East Arnhem	24,900
17. Iseki *et al.* (1994)	≥4 1982 ACR criteria by medical record review	Tokutei-Sikkan program applicants, author referral, medical association referral and dialysis registry	1972-1991	Japan	Okinawa	1,100,000
18. Hopkinson *et al.* (1994)	≥4 ACR criteria by patient interview	Hospital physician survey, immunology department registry, immunology lab, renal unit database, inpatient medical records and acute psychiatric admissions	1989-1990	United Kingdom	Nottingham	601,693
19. Grennan and Bossingham (1995)	≥4 1982 ACR criteria	Rheumatologist referral, Aboriginal health centres, hospital dermatology unit and outreach clinic	1993-1994	Australia	Northern Sydney and Queensland	45,305
20. Johnson *et al.* (1995)	≥4 ACR criteria by patient interview	National and private physician referral, lupus patient groups, rheumatology database, immunology laboratories and hospital discharge database	1991-1992	United Kingdom	Birmingham and Slihull districts	872,877
21. McCarty *et al.* (1995)	≥4 1982 ACR criteria by chart review	Rheumatologist referral and hospitals and university SLE databases	1985-1990	USA	Allegheny County	1,336,449
22. *Gourley et al.* (1997)	SLE diagnosis by chart review	Case records, physician referral, Lupus UK and laboratory records	1992-1993	Ireland	Northern Ireland	1,631,800
23. Voss *et al.* (1998)	≥4 1982 ACR criteria	Inpatient & outpatient registry, GP & specialist referral and university autoimmune test database	1980-1994	Denmark	Funen county	387,871
24. Uramoto *et al.* (1999)	≥4 1982 ACR criteria by medical records review	Community diagnostic retrieval system for SLE, ANA, LE cell and false-positive syphilis	1950-1992	USA	Rochester	not provided
25. Peschken *et al.* (2000)	≥4 1982 ACR criteria and diagnosed by physician	Specialist and family physician personal patient databases	1980-1996	Canada	Manitoba	1,100,295
26. Stahl-Hallengren *et al.* (2000)	Previously validated method	Hospital & primary healthcare registry, private clinic network, private physician referral and laboratory	1981-1991	Sweden	Lund and Orup districts	174,952
27. Chakravarty *et al.* (2000)	≥4 1997 ACR criteria by medical chart review	Hospital inpatient database	2000-2000	USA	California Pennsylvania	not provided not provided
28. Segasothy and Phillips (2001)	≥4 1997 ACR criteria	Medical records of patients with SLE or ANA titre ≥ 1:40	1990-1999	Australia	Central Australia	50,000

Study Author	Case definition	Sources for case ascertainment	Period	Country	Location	Population
29. Nossent (2001)	≥4 1982 ACR criteria validated by medical chart review	Inpatient database, outpatient registry, GP referral and national mortality database	1978-1996	Norway	Finnmark and Trom counties	224,403
30. Walsh *et al.* (2001)	≥3 1982 ACR criteria by chart review and patient examination	Community referral	1997-1997	USA	Nogales	19,489
31. Bossingham (2002)	≥4 1982 ACR criteria by medical records review	Local health care staff referral	1996-1998	Australia	Queensland	238,000
32. Vilar and El Sato (2002)	≥4 1982 ACR criteria	University hospital, public health network establishments, specialist referral and ANA laboratories	2000-2000	Brazil	Natal	709,422
33. Deligny *et al.* (2002)	≥4 1982 ACR criteria by medical record review/ patient assessment	Registry, physician referral, laboratory files, death registry and social security files	1990-1999	French West Indies	Martinique	381,427
34. Al-Arfaj *et al.* (2002)	≥4 1982 ACR criteria by patient assessment	Personal interview	1992-1992	Saudi Arabia	Al-Qaseem	10,372
35. Alamanos *et al.* (2003)	≥4 1982 ACR criteria confirmed by medical record review	Inpatients and outpatients referred to public and private rheumatology practices	1982-2001	Greece	Six northeast districts	488,435
36. Benucci *et al.* (2003)	Positive SLE screen, ≥4 1997 ACR criteria by patient assessment	Local GP's administering the Lupus Screening Questionnaire	2002-2002	Italy	Florence, Italy	32,521
37. Lopez *et al.* (2003)	≥4 ACR criteria by patient assessment	Centralized immunology laboratory database	1998-2002	Spain	Asturias	1,073,971
38. *Ward et al.* (2004)	Self-reported physician diagnosis and drug treatment	Patient self-report and drug treatment from NHANES III	1988-1994	USA	Country-wide	20,050
39. Naleway *et al.* (2005)	≥4 1982 ACR criteria by medical records review	Community clinic electronic SLE records	1991-2001	USA	Wisconsin	77,280
40. *Govoni et al.* (2006)	≥4 1982 ACR criteria by medical records review	Search for SLE (ICD-9: 710.0) within the hospital database and the national health care system database	1996-2002	Italy	Ferrara district	346,000
41. Nightingale *et al.* (2006)	4 ACR criteria, SLE in medical records or SLE drug prescription	General Practice Research Database, medical records and prescription records	1992-1998	United Kingdom	Country-wide	12,911,216 PY's
42. Bernatsky *et al.* (2007)	Case ascertainment algorithm	Billing codes, hospital discharge data and procedure code data from administrative databases	1994-2003	Canada	Quebec	7,500,000
43. Somers *et al.* (2007)	General Practitioner, SLE diagnostic codes	General Practice Research Database	1990-1999	United Kingdom	Country-wide	3,366,320 PY's

Appendix B: SLE studies estimating incidence

Study Author	Country	Location	Group	Yearly Incidence and 95% CI (/100,000)					
				Females		Males		Both Sexes	
1. Siegel *et al.* (1970)	USA	Jefferson County	Everyone	--	--	--	--	9.9	--
			African-American	28.8	--	0.3	--	--	--
			Caucasian	10.6	--	0.4	--	--	--
		New York City	Everyone	--	--	--	--	18.5	--
			African-American	76.5*	--	1.14*	--	--	--
			Caucasian	23.9*	--	0.27*	--	--	--
2. Fessel (1974)	USA	San Francisco	Everyone	--	--	--	--	7.6	--
3. Amor *et al.* (1983)	France	Country-wide	Everyone	--	--	--	--	0.1	--
6. Nived *et al.* (1983)	Sweden	Lund and Orup	Everyone	7.6	--	2.0	--	4.8	--
7. Hochberg (1985)	USA	Baltimore	Everyone	--	--	--	--	4.6*	--
	USA		African-American	10.5	--	2.1	--	--	--
	USA		Caucasian	3.9	--	0.45	--	--	--
8. Michet *et al.* (1985)	USA	Rochester	Everyone	2.5	[1.4, 3.6]	0.9	[0.2, 1.6]	1.8	[1.1, 2.5]
12. Gudmundsson and Steinsson (1990)	Iceland	Iceland	Everyone	5.8*	--	0.8*	--	3.3*	--
13. Jonsson *et al.* (1990)	Sweden	Southern	Everyone	5.4	--	1.1	--	4.0	[1.6, 6.4]
15. Nossent (1992)	Curacao	Country-wide	Afro-Caribbean	7.9	[2.3, 13.2]	1.1	[-0.9, 3.1]	4.6	[0.4, 8.8]
16. Anstey *et al.* (1993)	Australia	Darwin, Katherine & East Arnhem	Aboriginal	--	--	--	--	11.0	--
17. Iseki *et al.* (1994)	Japan	Okinawa	Everyone	4.2	[35.1, 48.8]	0.4	[0.3, 0.6]	0.9	--
18. Hopkinson *et al.* (1994)	United	Nottingham	Everyone	6.5*	[3.6, 9.5]	1.5	[0, 2.9]	4.0*	[2.4, 5.7]
20. Johnson *et al.* (1995)	United Kingdom	Birmingham and Slihull districts	Everyone	6.8*	[4.4, 9.2]	0.5*	[0.1, 1.7]	3.8*	[2.5, 5.1]
			Afro-Caribbean	22.8*	[6.9, 50.2]	--	--	--	--
			Asian	29.2*	[14.2, 51.0]	--	--	--	--
			Caucasian	4.5*	[2.7, 7.2]	--	--	--	--

Study Author	Country	Location	Group	Yearly Incidence and 95% CI (/100,000)					
				Females		Males		Both Sexes	
21. McCarty *et al.* (1995)	USA	Allegheny County	Everyone	--	--	--	--	2.4	[2.1, 2.8]
			African-American	9.2	[6.8, 12.5]	0.7	[0.0-2.0]	5.3	[3.9, 7.0]
			Caucasian	3.5	[2.9, 4.2]	0.4	[0.2, 0.7]	2.0	[1.7, 2.4]
23. Voss *et al.* (1998)	Denmark	Funen county	Everyone	--	--	--	--	2.5	[1.8, 3.3]
24. Uramoto *et al.* (1999)	USA	Rochester	Everyone	--	--	--	--	3.06	--
25. Peschken *et al.* (2000)	Canada	Manitoba	Everyone	--	--	--	--	--	--
			Aboriginal	--	--	--	--	2.0-7.4	--
			Caucasian	--	--	--	--	0.9-2.3	--
26. Stahl-Hallengren *et al.* (2000)	Sweden	Lund and Orup districts	Everyone	--	--	--	--	4.5	--
29. Nossent (2001)	Norway	Finnmark and Trom counties	Everyone	4.6	--	0.6	--	2.6	[1.9, 2.9]
32. Vilar and El Sato (2002)	Brazil	Natal	Everyone	14.1	[10.0, 19.3]	2.2	[0.7, 5.2]	8.7	[6.3, 11.7]
33. Deligny *et al.* (2002)	French	Martinique	Everyone	8.5	[7.2, 9.8]	0.7	[0.3, 1.1]	4.7	[2.5, 6.9]
35. Alamanos *et al.* (2003)	Greece	Six northeast districts	Everyone	3.7	[3.2, 4.2]	0.49	[0.45, 0.53]	1.9	[1.5, 2.3]
37. Lopez *et al.* (2003)	Spain	Asturias	Everyone	3.6	[2.7, 4.4]	0.5	[0.3, 0.8]	2.2	[1.8, 2.5]
39. Naleway *et al.* (2005)	USA	Wisconsin	Everyone	8.2	[5.5, 10.9]	1.9	[0.6, 3.3]	5.1	[3.6, 6.6]
40. Govoni *et al.* (2006)	Italy	Ferrara district	Everyone	--	--	--	--	1.9	[1.1, 2.7]
41. Nightingale *et al.* (2006)	United Kingdom	Country-wide	Everyone	5.3	[4.8, 5.9]	0.7	[0.5, 0.9]	3.0	[2.7, 3.3]
42. Bernatsky *et al.* (2007)	Canada	Quebec	Everyone	--	--	--	--	3.0	[2.6, 3.4]
43. Somers *et al.* (2007)	United Kingdom	Country-wide	Everyone	7.89*	[7.46, 8.31]	1.53*	[1.34, 1.71]	4.87	[4.48, 4.94]

*Incidence estimate is adjusted by a standard

Appendix C: SLE studies estimating prevalence

Study Author	Country	Location	Group	Prevalence and 95% CI (/100,000)					
				Females		Males		Both sexes	
1. Siegel *et al.* (1970)	USA	Jefferson County	Everyone	--	--	--	--	5.73	--
			African-American	17.9	--	0.0	--	--	--
			Caucasian	7.5	--	1.8	--	--	--
		New York City	Everyone	--	--	--	--	13.4	--
			African-American	53.7*	--	3.03*	--	--	--
			Caucasian	16.2*	--	2.94*	--	--	--
2. Fessel (1974)	USA	San Francisco	Everyone	--	--	--	--	50.8	--
4. Hart *et al.* (1983)	New Zealand	Auckland	Everyone	--	--	--	--	17.62*	--
			Caucasian	--	--	--	--	14.6*	--
			Polynesian	--	--	--	--	50.63*	--
5. Helve (1985)	Finland	Finland	Everyone	--	--	--	--	28	--
6. Nived *et al.* (1985)	Sweden	Lund and Orup	Everyone	64.8	--	11.7	--	39	[30, 48]
8. Michet *et al.* (1985)	USA	Rochester	Everyone	53.8	[27.4, 80.2]	19.0	[1.0, 37.0]	40.0	[23.5, 57.5]
9. Hochberg *(1987)*	United Kingdom	England and Wales	Everyone	12.5	--	--	--	6.5	--
10. Samanta *et al.* (1989)	United Kingdom	Leicaster	Asian	--	--	--	--	20	--
			Caucasian	--	--	--	--	40	--
11. Maskarinec and Katz (1989)	USA	Hawaii	Everyone	--	--	--	--	41.8	--
12. Gudmundsson and Steinsson (1990)	Iceland	Iceland	Everyone	62.0*	--	7.2*	--	35.9*	--
14. Boyer *et al.* (1991)	USA	Alaska	Aboriginal	--	--	--	--	112.2*	--
15. Nossent (1992)	Curacao	Country-wide	Afro-Caribbean	83.9	[65.8, 101.8]	8.5	[2.8, 14.2]	47.0	[34.1, 51.1]
16. Anstey *et al.* (1993)	Australia	Darwin, Katherine & East Arnhem	Aboriginal	100	--	5.2	--	52	--
17. Iseki *et al.* (1994)	Japan	Okinawa	Everyone	68.4	--	7	--	37.7	--

Study Author	Country	Location	Group	Prevalence and 95% CI (/100,000)					
				Females		Males		Both sexes	
18. Hopkinson *et al.* (1994)	United Kingdom	Nottingham	Everyone	--	--	--	--	24.7*	[20.7, 28.8]
			Afro-Caribbean	--	--	--	--	207.0*	[111, 302]
			Asian	--	--	--	--	48.8*	[10.5, 87.1]
			Caucasian	--	--	--	--	20.3*	[16.6, 24.0]
19. Grennan and Bossingham (1995)	Australia	Northern Queensland and Sydney	Aboriginal	--	--	--	--	44.2	--
20. Johnson *et al.* (1995)	United Kingdom	Birmingham and Slihull districts	Everyone	49.6*	[43.2, 56.1]	3.6*	[2.0, 6.0]	27.7*	[24.2, 31.2]
			Afro-Caribbean	197.2*	[162.5, 323.0]	6.4*	[0.1, 26.0]	111.8*	[80.8, 142.8]
			Asian	96.5*	[74.5, 118.6]	4.3*	0.6, 18.7]	46.7*	[31.5, 61.9]
			Caucasian	36.3*	[33.2, 39.3]	3.4*	[1.8, 5.9]	20.7*	[17.5, 24.0]
22. Gourley *et al.* (1997)	Ireland	Northern Ireland	Everyone	46.5	--	4.3	--	25.4	[22.1, 28.7]
23. Voss *et al.* (1998)	Denmark	Funen county	Everyone	37.9	[29.8, 47.5]	4.7	[2.2, 9.0]	21.7	[17.3, 26.8]
24. Uramoto *et al.* (1999)	USA	Rochester	Everyone	--	--	--	--	122.0*	--
25. Peschken *et al.* (2000)	Canada	Manitoba	Everyone	--	--	--	--	22.1	[13.2, 32.4]
			Aboriginal	--	--	--	--	33.4	[22.7, 41.8]
			Caucasian	--	--	--	--	20.6	[12.4, 30.8]
26. Stahl-Hallengren *et al.* (2000)	Sweden	Lund and Orup districts	Everyone	--	--	--	--	68	--
27. Chakravarty *et al.* (2000)	USA	California	Everyone	184.2	[181.4, 187.0]	25.5	[24.5, 26.6]	107.7	[106.1, 109.2]
		Pennsylvania	Everyone	253	[248.3, 257.7]	38.7	[36.8, 40.7]	149.5	[146.9, 152.2]
28. Segasothy and Phillips (2001)	Australia	Central Australia	Aboriginal	--	--	--	--	19.3	--
			Caucasian	--	--	--	--	73.5	--
29. Nossent (1992)	Norway	Finnmark and Trom counties	Everyone	89.3*	[78.9,100.2]	9.7*	[6.9, 12.6]	49.7*	[44.3, 55]
30. Walsh *et al.* (2001)	USA	Nogales	Everyone	--	--	--	--	94.0	--
31. Bossingham (2002)	Australia	Queensland	Everyone	--	--	--	--	45.3	--
			Aboriginal	--	--	--	--	92.8	--
33. Deligny *et al.* (2002)	French West Indies	Martinique	Everyone	115	[100, 130]	9.2	[4.8, 13.6]	64.2	[56.2, 72.2]

Study Author	Country	Location	Group	Prevalence and 95% CI (/100,000)					
				Females		Males		Both sexes	
34. Al-Arfaj *et al.* (2002)	Saudi Arabia	Al-Qaseem	Everyone	--	--	--	--	19.28	--
35. Alamanos *et al.* (2003)	Greece	Six northeast districts	Everyone	67.3*	[64.4, 71.5]	9.1*	[5.8, 12.4]	38.1*	[36.3, 39.9]
36. Benucci *et al.* (2003)	Italy	Florence, Italy	Everyone	--	--	--	--	71	[49, 92]
37. Lopez *et al.* (2003)	Spain	Asturias	Everyone	57.9	[51.6, 64.2]	8.3	[5.8, 10.8]	34.1	[30.6, 37.6]
38. Ward *et al.* (2004)	USA	Country-wide	Everyone	100.0	[19.8, 179.3]	3.4	[0, 10.2]	53.6	[12.2, 95.0]
39. Naleway *et al.* (2005)	USA	Wisconsin	Everyone	131.5	[95.5, 167.5]	24.8	[9.4, 40.2]	78.5*	[59.0, 98.0]
40. Govoni *et al.* (2006)	Italy	Ferrara district	Everyone	100.1	--	12	--	57.9	--
41. Nightingale *et al.* (2006)	United Kingdom	Country-wide	Everyone	70.8*	[65.1, 76.6]	10.0*	[7.8, 12.2]	40.7	[37.6, 43.8]
42. Bernatsky *et al.* (2007)	Canada	Quebec	Everyone	45.3*	[37.6, 53.0]	3.7*	[1.5, 5.9]	44.7	[37.4, 54.7]

*Prevalence estimate is adjusted by a standard