Rheumatol Int
DOI 10.1007/s00296-010-1591-2

SHORT COMMUNICATION

Surveillance of systemic autoimmune rheumatic diseases using administrative data

S. Bernatsky · L. Lix · J. G. Hanly · M. Hudson · E. Badley · C. Peschken · C. A. Pineau · A. E. Clarke · P. R. Fortin · M. Smith · P. Bélisle · C. Lagace · L. Bergeron · L. Joseph

Received: 20 April 2010 / Accepted: 14 July 2010

Abstract There is growing interest in developing tools and methods for the surveillance of chronic rheumatic diseases, using existing resources such as administrative health databases. To illustrate how this might work, we used population-based administrative data to estimate and compare the prevalence of systemic autoimmune rheumatic diseases (SARDs) across three Canadian provinces, assessing for regional differences and the effects of demographic factors. Cases of SARDs (systemic lupus erythematosus, scleroderma, primary Sjogren's, polymyositis/dermatomyositis) were ascertained from provincial physician billing and hospitalization data. We combined information from three case definitions, using hierarchical Bayesian latent class regression models that account for the imperfect nature of each case definition. Using methods that account for the imperfect nature of both billing and hospitalization databases, we estimated the over-all prevalence of SARDs to be approximately 2–3 cases per 1,000 residents. Stratified prevalence estimates suggested similar demographic trends across provinces (i.e. greater prevalence in females-versus-males, and in persons of older age). The prevalence in older females approached or exceeded 1 in

S. Bernatsky (✉) · A. E. Clarke · P. Bélisle · L. Joseph
Division of Clinical Epidemiology,
Research Institute of the McGill University Health Centre
(MUHC), 687 Pine Avenue West, V-Building,
V2.09, Montreal, QC H3A 1A1, Canada
e-mail: sasha.bernatsky@mail.mcgill.ca

S. Bernatsky · C. A. Pineau
Division of Rheumatology, MUHC, Montreal, PQ, Canada

L. Lix
School of Public Health, University of Saskatchewan,
Saskatoon, SK, Canada

J. G. Hanly
Division of Rheumatology, Department of Medicine
and Department of Pathology, Dalhousie University
and Queen Elizabeth II Health Sciences Centre,
Halifax, NS, Canada

M. Hudson
Division of Rheumatology, Jewish General Hospital,
Montreal, PQ, Canada

E. Badley
Dalla Lana School of Public Health, University of Toronto,
Toronto, ON, Canada

C. Peschken
Department of Medicine, University of Manitoba,
Winnipeg, MB, Canada

A. E. Clarke
Division of Clinical Immunology and Allergy,
MUHC, Montreal, PQ, Canada

P. R. Fortin
Toronto Western Hospital, University Health Network,
and University of Toronto, Toronto, ON, Canada

M. Smith
Manitoba Centre for Health Policy, University of Manitoba,
Winnipeg, MB, Canada

C. Lagace
Public Health Agency of Canada, Ottawa, ON, Canada

L. Bergeron
Canadian Arthritis Patient Alliance, Toronto, ON, Canada

L. Joseph
Department of Epidemiology, Biostatistics & Occupational
Health, McGill University, Montreal, PQ, Canada

Published online: 28 July 2010

100, which may reflect the high burden of primary Sjogren's syndrome in this group. Adjusting for demographics, there was a greater prevalence in urban-versus-rural settings. In our work, prevalence estimates had good face validity and provided useful information about potential regional and demographic variations. Our results suggest that surveillance of some rheumatic diseases using administrative data may indeed be feasible. Our work highlights the usefulness of using multiple data sources, adjusting for the error in each.

Keywords Systemic autoimmune rheumatic diseases · Surveillance · Administrative databases · Prevalence

Introduction

Systemic autoimmune rheumatic diseases (SARDs) are complex chronic disorders, associated with high morbidity and mortality. In the developed world, there is increasing interest in monitoring the prevalence of rheumatic disease in our aging populations [1], particularly as the pool of specialists (e.g. rheumatologists) declines [2]. Accurate estimates of rates greatly improve our understanding of the clinical significance of these diseases and their public health impact. This is important because SARDs (including systemic lupus erythematosus (SLE), scleroderma, and polymyositis-dermatomyositis, for example) appear to be associated with significant morbidity and mortality. Direct medical costs are significant [3–5]; in some subsets of patients with SARDs (such as those with loss of renal function as a consequence of their disease), this cost may be in excess of $26,000 Canadian (16,750 Euros) per patient year [6]. Indirect costs, such as lost productivity, can average over $14,000 Canadian (8,700 Euros) per patient per year [7].

Decision-makers in Canada are interested in developing tools and methods for the surveillance of chronic diseases, using existing resources such as administrative health databases [8]. In order to support and inform the development of tools and methods for the surveillance of rheumatic conditions, such as SARDs, we used population-based administrative data to estimate SARD prevalence across three provinces in Canada, a country with essentially complete healthcare coverage under a single payer system. In addition, we provide sensitivity estimates of different case definition approaches and illustrate how regional comparisons might be used.

Materials and methods

Our research was approved by the provincial and academic ethics review boards in Quebec, Nova Scotia and Manitoba, and the procedures followed were in accordance with the Helsinki Declaration of 1975, as revised in 1983.

We relied on physician billing and hospitalization databases covering all residents of Manitoba (1.1 million), Quebec (7.5 million), and Nova Scotia (913,000). The billing databases document virtually all physician services (with one diagnostic code, assigned by the physician, for each visit), and all in-patient stays (with discharge diagnoses for each hospitalization). In these databases, SARD diagnoses are captured under International Classification of Diseases (ICD-9) code 710. This includes systemic lupus erythematosus (SLE), Sjogrens, scleroderma (diffuse and limited), polymyositis, and dermatomyositis.

We used three case definitions, one based on hospitalization data, the other two based on physician billing data. Within hospitalization data, we defined a SARD case on the basis of any discharge diagnosis with a relevant ICD code. Of our two billing-code definitions, the first one defined a case according to an algorithm requiring at least 2 physician visits for a SARD, at least 2 months apart, but within a 2-year span. The second approach using billing data defined cases as those individuals with at least one SARD billing code contributed by a rheumatologist. Some individuals would be detected by one or more, but not necessarily all, of these three definitions. However, for prevalence estimates, an individual could be counted only once. The prevalence estimates included all cases identified (that is, anyone who met at least one of the three definitions) who remained provincial residents as of December 31st of the last year of the study period (this was 1995–2003 for Manitoba, 1994–2003 for Quebec, and 1995–2004 for Nova Scotia).

We adjusted for the imperfect sensitivity and specificity of the data, using previously developed methods, which do not assume the existence of a gold standard [9]. These methods allow estimation not only of disease prevalence, but also of the sensitivity and specificity of each of the three different case definitions. Without a feasible 'gold standard', the sensitivity and specificity of a given case definition cannot be calculated directly. This is because the true disease state for each subject (who may be positive according to one or more case definitions and negative according to other case definitions) is unknown, or 'latent' [10]. However, multiple case definitions can each provide some information about the case status of subjects, allowing disease status (and prevalence) to be estimated probabilistically. To do so, we used a Bayesian approach to latent class regression, which further allowed us to combine divergent results (e.g. individuals who are defined as a case by at least one definition but not by all) to produce sensitivity estimates for each of the case definitions. These sensitivity estimates are relative to the true disease status of individuals, which is not directly observed but is itself estimated by our models.

Bayesian methods [11] are based on the idea that unknown values for a parameter have probability distributions. A Bayesian analysis begins with prior probability distributions for the unknown parameters of interest. The prior distributions summarize all relevant previous information about the parameters of interest. A prior distribution is then updated by new data. This results in a posterior distribution, which represents what one should now believe about the parameter values, given the initial background and the new data. The methodology is underpinned by Bayes' theorem a mathematical rule for updating prior beliefs in the light of new data.

Since two of our ascertainment definitions were derived from a similar source (physician billing claims), our model also had to adjust for the possible correlation between these two definitions. In this setting, we could only estimate the parameters of interest (SARD prevalence and the sensitivities of the three different case definitions) if we used informative prior distributions for at least some of the specificities [12]. In earlier studies of SARDs using administrative databases [13, 14], the specificities of all methods of case ascertainment were very high, generally greater than 98%. Thus, for our primary analyses we set informative beta (248.3, 1.65) prior distributions for the specificities of our two billing data case ascertainment approaches. This prior corresponds to specificities of 99% (95% credible interval 98, 100). (In sensitivity analyses, we set alternative specificity priors corresponding to 94% (88 to 100). As the results using these different sets of prior inputs were similar, we only reported results from the primary analyses.) We employed low information (diffuse) prior distributions for all other parameters (e.g. for prevalence ß[1, 1]), both reflecting the fact that there is little existing information about them, and that we wanted to use the data to estimate these parameters to the extent possible.

Disease prevalence (and potentially, the sensitivity of different approaches to detecting disease) generally varies according to demographics and other factors. A 'hierarchical' model allowed us to account for these differences. Our use of latent class Bayesian regression models in this context has been described previously [12, 15, 16], though with different case definitions in other diseases. Briefly, the levels of our hierarchical model accounted for (1) Population sampling variability (assigned a binomial distribution) and misclassification error, adjusting for both false positives and false negatives; (2) Variations in disease prevalence related to patient demographics (age, sex, and urban-versus-rural residence), input as a logistic regression model on the binomial probabilities from the first level of the model; (3) Variations in case ascertainment sensitivity according to the same patient demographics, input as a distinct parameter for the sensitivity of each case definition. Urban-versus-rural residence was defined on the basis of postal-code data (urban areas defined on the basis of Census Metropolitan Area classifications) [17]. We handled potential conditional dependence (i.e., between the two different case definitions which were both based on billing data) with a covariance term, similar to the approach of Dendukuri and Joseph 2001 [12], based on ideas from Vacek [18]. The statistical software used was WinBUGS version 1.4.3 (MRC Biostatistics Unit, Cambridge, UK).

Results

Table 1 provides SARD prevalence estimates from the Bayesian latent class hierarchical models, for each province, both for age and sex-specific rates, and over-all. The total prevalence (Table 1c) in each province was between 2 and 3 cases/1,000; SARDs prevalence was marginally higher in Manitoba compared to Quebec. The highest prevalence was seen among individuals aged ≥45 (especially women). There appeared to be a trend toward a markedly higher female-to-male ratio in Manitoba (approximately 8:1, based on the prevalence rates in the table) compared to Nova Scotia (4:1) or Quebec (5:1). As can be seen in the table, greater prevalence in urban-versus-rural settings was evident in Nova Scotia and Quebec, but not Manitoba.

Our models estimated sensitivity for each of the three case definitions (rheumatology billing, two-code physician billing, and hospital diagnosis) within each province (Fig. 1). The sensitivity of the case definition based on two or more SARD codes in billing data was at best between 70 and 90%, and lower (50–70%) in older individuals. For rheumatology billing data, the sensitivity estimates were about 50–70% in younger individuals and somewhat lower (40–50%) in older individuals. Rheumatology billing data sensitivity estimates tended to be higher for urban (versus rural) residents. Hospitalization data was the least sensitive, across provinces and demographics. This was particularly true for Nova Scotia. In all provinces, it seemed that hospitalization data were more sensitive for the detection of males versus females SARD cases.

Discussion

Existing North American estimates suggest SLE prevalence at about 54/100,000, scleroderma at 28/100,000[19], and inflammatory myopathies (polymyositis/dermatomyositis) at 21.5/100,000[9], with variable estimates for Sjogren's syndrome, which may be >156/100,000[20]. Combining these estimates (some of which are derived from small,

Table 1 Systemic autoimmune rheumatic disease (SARD) prevalence estimates (Bayesian latent class hierarchical modeling): prevalence estimates (95% credible interval) per 1,000

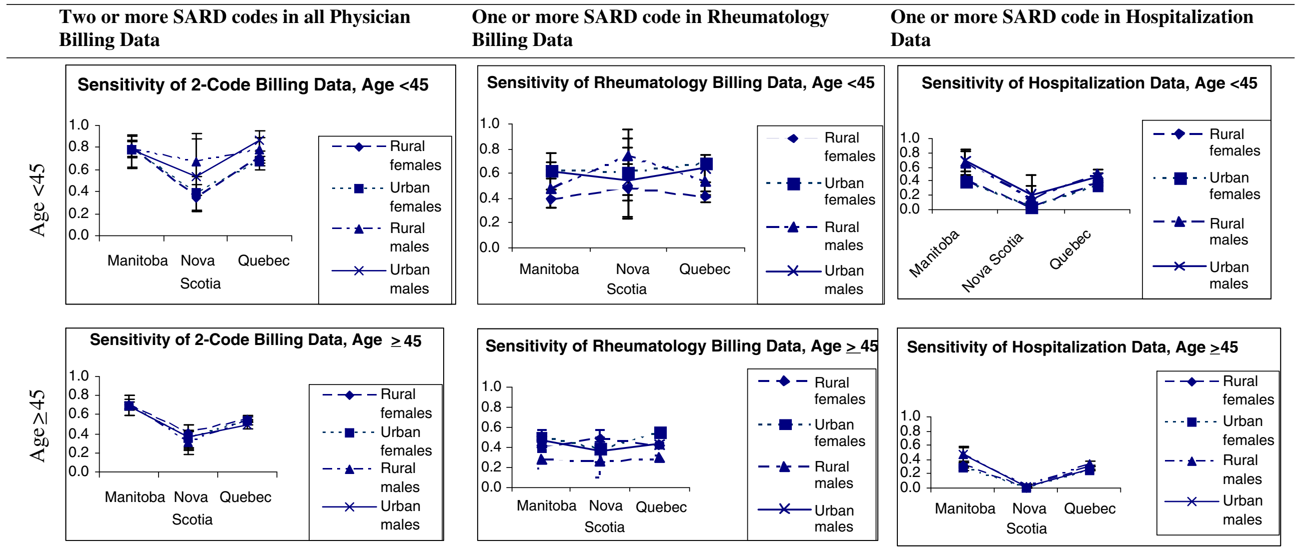
	Female		Male	
	Rural	Urban	Rural	Urban
a. Residents aged <45				
Manitoba	2.0 (1.7–2.3)	2.3 (2.0–2.6)	0.2 (0.1–0.2)	0.2 (0.2–0.3)
Nova Scotia	0.9 (0.6–1.4)	1.1 (0.8–1.7)	0.2 (0.1–0.3)	0.2 (0.1–0.4)
Quebec	1.0 (0.9–1.1)	1.9 (1.8–2.1)	0.1 (0.1–0.2)	0.2 (0.2–0.3)
b. Residents aged ≥45				
Manitoba	12.5 (11.6–13.4)	14.7 (13.9–15.5)	2.7 (2.3–3.2)	2.8 (2.5–3.4)
Nova Scotia	6.9 (5.5–10.4)	11.5 (9.2–16.7)	1.7 (1.1–2.6)	3.2 (2.3–4.8)
Quebec	5.6 (5.3–5.9)	9.4 (9.1–9.8)	1.4 (1.3–1.6)	2.3 (2.1–2.6)

	Females*	Males**	Total
c. Over-all prevalence			
Manitoba	6.9 (6.6–7.2)	1.2 (1.1–1.3)	4.1 (3.9–4.3)
Nova Scotia	4.2 (3.6–5.7)	1.0 (0.8–1.4)	2.7 (2.3–3.4)
Quebec	4.2 (4.1–4.4)	0.8 (0.8–0.9)	2.6 (2.5–2.7)

Credible intervals represent the values between which there is a 95% probability of containing the parameter of interest, given the data & prior information input

* Number of female SARD cases: Manitoba 2804, Nova Scotia 1373, Quebec 14,551

** Number of male SARD cases: Manitoba 464, Nova Scotia 314, Quebec 3436



Our models estimated sensitivity estimates for each of the 3 case definitions (rheumatology billing, 2-code physician billing, and hospital diagnosis), based on the total number of SARDs cases identified from all sources, and accounting for error (both in over-ascertainment and under-ascertainment) each. Error bars represent Bayesian credible intervals, the values between which there is a 95% probability of containing the parameter of interest, given the data & prior information input.

Fig. 1 Sensitivity of different systemic autoimmune rheumatic disease (SARD) case ascertainment methods (estimated from Bayesian hierarchical latent class models)

defined population survey sources) suggests an over-all prevalence of SARDs that is very close to our figures (between 2 and 3/1,000). This means SARDs are as common as inflammatory bowel disease (also about 2–3 cases/1,000) [21] in Canada.

Stratified prevalence estimates suggested similar demographic trends across provinces (i.e. greater prevalence in females-versus-males and with age ≥45). The prevalence in older females approached or exceeded 1 in 100 (Table 1), in large part likely due to primary Sjogren's syndrome, which has been shown to affect up to 1% of older females[3]. We noted a greater prevalence in urban-versus-rural settings, possibly due to sick rural patients migrating to urban areas, where more tertiary care is located. This was evident in Nova Scotia and Quebec, but not Manitoba, a phenomenon not easy to explain, since it might be due to a complex combination of geographic and socio-demographic factors, including patient preferences, culture, and regional distribution of resources. Also, the concept of urban-versus-rural area itself is not easily captured [22], and a more sophisticated treatment of this variable might be useful in future studies.

Administrative databases are not always complete or accurate. Physician billing databases in Canada may have limited sensitivity for ascertainment of chronic diseases, since generally only one diagnostic code per visit is allowed. Patients may have multiple co-morbidities, and theoretically, these could take precedence as the billing diagnosis (instead of a chronic rheumatic condition). Our methods would have underestimated SARD prevalence if, for example, some patients did not receive lupus-related care over our study period. Presumably, this might result in the under-ascertainment primarily of milder cases.

In addition to missing some true cases, any ascertainment method will also misclassify some persons if sources like medical record or classification criteria are considered the true gold standard. Preliminary work by our group to determine the accuracy of SARD diagnoses from administrative data (comparing these to medical records) suggests that the majority of individuals thus identified do have some type of SARD, although there is some misclassification between categories (SLE vs. scleroderma, for example). We note that model-based approaches to case definition from administrative data have been used recently by others to produce prevalence estimates for osteoporosis [23] and asthma [24.]

What real-life applications do our results have, for stakeholders who want to improve outcomes in rheumatic disease? One might begin by considering the slightly increased prevalence of SARDs in Manitoba versus other provinces. This could generate questions both in terms of determining what might drive this (for example, the high Aboriginal population in this province) as well as whether there are adequate resources to deal with this higher prevalence. There are trends for different physician-to-population ratios being lower in Manitoba when compared to elsewhere [25], though available data do not account for part-time versus full-time positions or academic versus clinical practice.

It is moreover interesting that our data appear to suggest differences in how patients present in a province like Nova Scotia versus elsewhere. In Nova Scotia, patients appeared much less likely to be identified from hospitalization databases. At this province's academic rheumatology centre, all rheumatology referrals are reviewed and triaged, and SARDs are included as a priority in terms of wait times to be seen. This innovation may create better access to rheumatology, which has the potential benefit of optimizing care and preventing hospitalizations.

Decision-makers, including public health care officials, are looking to administrative databases as a means of chronic disease surveillance. Our results suggest that such surveillance of some rheumatic diseases may indeed be feasible and useful. Our work highlights the usefulness of using multiple data sources, adjusting for the error in each. Yet, as is highlighted in the recent paper by Yiannakoulias et al. [26], there are many considerations in the use of administrative data for public health surveillance that require ongoing research. For one, the right geographic units need to be used, but this may depend on multiple factors, such as the nature of the disease under study, the question of interest, and so on. We are currently exploring more multi-level analyses in this regard.

Acknowledgments This study was funded by the Canadian Institutes of Health Research (CIHR). Dr. Sasha Bernatsky is a Canadian Arthritis Network Scholar and is supported by the CIHR, the Fonds de la Recherche en Santé du Québec (FRSQ) and the McGill University Health Centre (MUHC) Research Institute and Department of Medicine. Dr. Christian Pineau is supported by the MUHC Research Institute and Department of Medicine. Drs. Ann Clarke and Lawrence Joseph are FRSQ National Scholars. Dr. Marie Hudson holds a Junior Investigator CIHR career award. The authors are indebted to Manitoba Health and Healthy Living for the provision of data. The results and conclusions are those of the authors, and no official endorsement by Manitoba Health and Healthy Living is intended or should be inferred.

References

1. The Arthritis Society. Highlights from the first ever Canadian consensus conference on Systemic Autoimmune Rheumatic Diseases (SARD) Dec 20, 2007. Available at: http://www.arthritis.ca/look%20at%20research/sard/default.asp?s=1. Accessed 16 May 2009
2. Hanly JG (2001) Manpower in Canadian academic rheumatology units: current status and future trends. Canadian Council of Academic Rheumatologists. Rheumatol 28:1944–1951
3. Callaghan R, Prabu A, Allan RB et al (2007) Direct healthcare costs and predictors of costs in patients with primary Sjogren's syndrome. Rheumatology 46:105–111
4. Sutcliffe N, Clarke AE, Taylor R et al (2001) Total costs and predictors of costs in patients with systemic lupus erythematosus. Rheumatology (Oxford) 40:37–47
5. Bernatsky S, Hudson M, Panopalis P et al (2009) The cost of systemic sclerosis. Arthritis Rheum 61:119–123
6. Clarke AE, Panopalis P, Petri M et al (2008) SLE patients with renal damage incur higher health care costs. Rheumatology (Oxford) 47:329–333
7. Panopalis P, Petri M, Manzi S et al (2007) The systemic lupus erythematosus Tri-Nation study: cumulative indirect costs. Arthritis Rheum 57:64–70

8. Wilchesky M, Tamblyn RM, Huang A (2004) Validation of diagnostic codes within medical services claims. J Clin Epidemiol 57:131–141
9. Bernatsky S, Joseph L, Pineau CA et al (2008) Estimating the prevalence of polymyositis and dermatomyositis from administrative data: age, sex, and regional differences. Ann Rheum Dis 68:1192–1196
10. Joseph L, Gyorkos T, Coupal L (1995) Bayesian estimation of disease prevalence and the prevalence of diagnostic tests in the absence of a gold standard. Am J Epidemiol 141:263–272
11. Ashby D (2006) Bayesian statistics in medicine: a 25 year review. Stat Med 25:3589–3631
12. Dendukuri N, Joseph L (2001) Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. Biometrics 57:158–167
13. Losina E, Barrett J, Baron JA et al (2003) Accuracy of Medicare claims data for rheumatologic diagnoses in total hip replacement recipients. J Clin Epidemiol 56:515–519
14. Bernatsky S, Joseph L, Pineau CA et al (2007) A population-based assessment of systemic lupus erythematosus incidence and prevalence–results and implications of using administrative data for epidemiological studies. Rheumatology (Oxford) 46:1814–1818
15. Bernatsky S, Joseph L, Belisle P et al (2005) Bayesian modelling of imperfect ascertainment methods in cancer studies. Stat Med 24:2365–2379
16. Ladouceur M, Rahme E, Pineau CA et al (2007) Robustness of prevalence estimates derived from misclassified data from administrative databases. Biometrics 63:272–279
17. Statistics Canada. Geographic Units: Census Metropolitan Area (CMA) and Census Agglomeration (CA) December 17, 2002. Available at: http://www12.statcan.ca/english/census01/Products/Reference/dict/geo009.htm. Accessed 16 May 2009
18. Vacek PM (1985) The effect of conditional dependence on the evaluation of diagnostic tests. Biometrics 41:959–968
19. Helmick CG, Felson DT, Lawrence RC et al (2008) Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part I. Arthritis Rheum 58:15–25
20. Kabasakal Y, Kitapcioglu G, Turk T et al (2006) The prevalence of Sjogren's syndrome in adult women. Scand J Rheumatol 35:379–383
21. Bernstein CN, Wajda A, Svenson LW et al (2006) The epidemiology of inflammatory bowel disease in Canada: a population-based study. Am J Gastroenterol 101:1559–1568
22. Rourke J (2007) In search of a definition of "rural". Can J Rural Med 2:113–115
23. Lix LM, Yogendran MS, Leslie WD et al (2008) Using multiple data features improved the validity of osteoporosis case ascertainment from administrative databases. J Clin Epidemiol 61:1250–1260
24. Prosser RJ, Carleton BC, Smith MA (2008) Identifying persons with treated asthma using administrative data via latent class modelling. Health Serv Res 43:733–754
25. Statistical information on Canadian physicians 2009. Available at: http://www.cma.ca/index.cfm/ci_id/16959/la_id/1.htm
26. Yiannakoulias N, Svenson LW, Schopflocher DP (2009) An integrated framework for the geographic surveillance of chronic disease. Int J Health Geogr 8:69