Open Access Rheumatology: Research and Reviews

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

The use of administrative health care databases to identify patients with rheumatoid arthritis

This article was published in the following Dove Press journal:
Open Access Rheumatology: Research and Reviews
6 November 2015
Number of times this article has been viewed

John G Hanly[1,2]
Kara Thompson[3]
Chris Skedgel[4]

[1]Division of Rheumatology, Department of Medicine, [2]Department of Pathology, [3]Department of Medicine, Queen Elizabeth II Health Sciences Centre, Dalhousie University, [4]Atlantic Clinical Cancer Research Unit, Capital Health, Halifax, Nova Scotia, Canada

Objective: To validate and compare the decision rules to identify rheumatoid arthritis (RA) in administrative databases.

Methods: A study was performed using administrative health care data from a population of 1 million people who had access to universal health care. Information was available on hospital discharge abstracts and physician billings. RA cases in health administrative databases were matched 1:4 by age and sex to randomly selected controls without inflammatory arthritis. Seven case definitions were applied to identify RA cases in the health administrative data, and their performance was compared with the diagnosis by a rheumatologist. The validation study was conducted on a sample of individuals with administrative data who received a rheumatologist consultation at the Arthritis Center of Nova Scotia.

Results: We identified 535 RA cases and 2,140 non-RA, noninflammatory arthritis controls. Using the rheumatologist's diagnosis as the gold standard, the overall accuracy of the case definitions for RA cases varied between 68.9% and 82.9% with a kappa statistic between 0.26 and 0.53. The sensitivity and specificity varied from 20.7% to 94.8% and 62.5% to 98.5%, respectively. In a reference population of 1 million, the estimated annual number of incident cases of RA was between 176 and 1,610 and the annual number of prevalent cases was between 1,384 and 5,722.

Conclusion: The accuracy of case definitions for the identification of RA cases from rheumatology clinics using administrative health care databases is variable when compared to a rheumatologist's assessment. This should be considered when comparing results across studies. This variability may also be used as an advantage in different study designs, depending on the relative importance of sensitivity and specificity for identifying the population of interest to the research question.

Keywords: inflammatory arthritis, case definitions, incidence, prevalence, population health

Introduction

The study of chronic disease at a population health level allows estimation of overall disease frequency, health care utilization, and associated costs.[1] Such information is helpful to determine the most appropriate allocation of current and future health care resources. Rheumatoid arthritis (RA) is one of the most common chronic autoimmune rheumatic diseases.[2] It primarily affects young- and middle-aged adults and has serious individual and societal consequences if not diagnosed and treated early in the disease course.[3] Population estimates of this disease provide a basis for estimating current and increasing health care needs in the future.

Administrative health care databases that contain information on physician billings, hospital visits, and medications are frequently used in the study of population health.[4]

Correspondence: John G Hanly
Division of Rheumatology, Nova Scotia Rehabilitation Centre, 2nd Floor, 1341 Summer Street, Halifax, Nova Scotia, Canada B3H 4K4
Tel +1 902 473 3818
Fax +1 902 473 7019
Email john.hanly@cdha.nshealth.ca

To determine the diagnostic accuracy of this approach, various case definitions utilizing different combinations of variables have been examined.[5,6] However, the optimal methodology remains unclear. Our aim was to examine and validate a variety of case definitions that can be applied to the administrative databases to identify patients with RA.

Methods

Study populations and controls

Nova Scotia is a Canadian province of ~1 million inhabitants. There are 2,458 physicians in Nova Scotia, of which 1,227 work in primary care, 193 are general internists, and eleven are adult rheumatologists. Health care services, including acute hospitalizations and ambulatory physician visits, are universally provided as specified under the Canada Health Act. The data were derived from information on Nova Scotia residents who were enrolled in the Medical Services Insurance program between April 1997 and March 2011. This excludes status native Canadians and members of the Canadian armed forces. The validation study was conducted on a further subset of individuals who had received at least one rheumatologist consultation at the Arthritis Center of Nova Scotia.

The data were obtained from existing databases accessed through Health Data Nova Scotia in the Department of Community Health and Epidemiology at Dalhousie University in Halifax, Nova Scotia, Canada. Within this unit, there are secure research-computing facilities on site, and access to data is governed by Health Data Nova Scotia Data Access Guidelines and Procedures. The study protocol was reviewed and approved by the Capital Health Research Ethics Board. Informed consent from individual patients was not required as the study utilized secondary administrative data.

Incident cases of RA in the administrative databases were defined as those without a physician billing for the same diagnosis in the preceding 5 years.[7] Prevalent cases included both incident and nonincident cases. The average annual incidence and prevalence rate was determined between April 2002 and March 2011. Patients with RA were matched 1:4 by age and sex to a control cohort of patients enrolled in the same databases but without a diagnosis of RA or other inflammatory arthritis (IA). The latter were excluded from the control cohort as we wished to distinguish RA from noninflammatory musculoskeletal conditions.

Data collection

Individual level data were obtained. Computerized claims were linked by encrypted health card number to the Canadian Institute of Health Information hospital discharge abstracts and Medical Services Insurance physician billings.

Comparison of administrative case definitions for RA with a rheumatologist's diagnosis

Validation of RA case definitions derived from the administrative data sets utilized information from rheumatology consultations at the Arthritis Center of Nova Scotia. In this analysis, the rheumatologist's diagnosis was taken as the gold standard for both RA cases (case-rheum) and controls (control-rheum). The data for the validation exercise used patient's information in the administrative data set and information from rheumatology consultations at the Arthritis Center identified through the clinic registration system. In this way, the analysis was restricted to patients who were in both the health administrative data set and the medical records of the Arthritis Center. The process maintained patient's confidentiality in compliance with Nova Scotia Provincial privacy legislation on use of health administrative data. The Arthritis Center is a regional resource with eight attending staff rheumatologists, and the majority of referrals are received from primary care physicians. Although most of the referred patients reside in the Capital Health District of Nova Scotia, which has 40% of the population of the province, a substantial number of patients are also referred from each of the other eight health districts.

Utilizing the validation data set described earlier, individuals were identified in the administrative data set who fulfilled one or more of the case definitions for RA. Four age- and sex-matched controls were identified for each case and designated as case-admin and control-admin. Scrambled unique identifiers (provincial health card numbers) were used to identify cases and controls, which were present in both the administrative data set and the Arthritis Center's record of clinic visits. The unique identifiers were then descrambled by an authorized third party to permit identification and review of the medical records in the Arthritis Center. To determine the validity of the diagnosis of RA in the administrative data, the individuals identified using each of the seven case definitions for RA and matched controls (case-admin and control-admin) were cross-referenced with the clinical diagnosis by a rheumatologist (case-rheum and control-rheum), arising from one or more ambulatory rheumatology assessments at the Arthritis Center. The clinical diagnosis derived through chart review for both cases and

controls was taken as the gold standard, and the review was done without knowledge of the administrative data for each of the seven case definitions. The estimated frequency of RA cases in the Arthritis Center's ambulatory clinics between 2000 and 2012 was 2,692/25,888 (10.4%) of the total clinic population.

Case definitions for identification of RA cases and validation

The following seven individual case definitions were used to identify the cases of RA in the administrative databases:

#1 MacLean:[8] Two physician visits for RA at least 2 months apart.

#2 MacLean/Lacaille:[7] MacLean algorithm with Lacaille variation, ie, excluding individuals with at least two visits, at least 2 months apart, subsequent to the second RA visit, with two identical diagnoses of other IAs and connective tissue diseases (psoriatic arthritis, ankylosing spondylitis, and other spondyloarthropathies, SLE, scleroderma, Sjogren's syndrome, dermatomyositis, polymyositis, other connective tissue diseases, and primary systemic vasculitis) and excluding those where a diagnosis of RA by a nonrheumatologist was not confirmed if/when the individual saw a rheumatologist.

#3 Shipton-like: Three RA diagnostic billing codes, over any time period, rather than in 3 consecutive years as described by Shipton et al.[9]

#4 Hospitalization: At least one hospitalization where RA was in the diagnostic codes.

#5 Rheumatologist: At least one RA code contributed by a rheumatologist.

#6 Combination: MacLean-like algorithm (two nonrheumatology physician visits for RA at least 2 months apart, within a 2-year period) or at least one RA code contributed by a rheumatologist or at least one hospitalization where RA was in the diagnostic codes and Lacaille variation, ie, excluding individuals with at least two visits, at least 2 months apart, subsequent to the second visit, with two identical diagnoses of other IAs and connective tissue diseases (psoriatic arthritis, ankylosing spondylitis, and other spondyloarthropathies, SLE, scleroderma, Sjogren's syndrome, dermatomyositis, polymyositis, other connective tissue diseases, and primary systemic vasculitis) and excluding those where a diagnosis of RA by a nonrheumatologist was not confirmed if/when the individual saw a rheumatologist.

#7 Single admin: Any single diagnostic code for RA.

The following *International Classification of Diseases, Ninth Edition (ICD-9) and International Classification of Diseases, Tenth Edition (ICD-10)* diagnostic codes were used:

RA
(*ICD-9*: 714.0, 714.1, and 714.2. *ICD-10:* MO5–MO5.9, MO6.0, MO6.8, and MO6.9).

Systemic lupus erythematosus
(*ICD-9:* 710.0. *ICD-10: ICD-10*: M 32, M32.1, M32.8, and M32.9).

Psoriatic arthritis
(*ICD-9*: 696.0. *ICD-10*: L40.5).

Ankylosing spondylitis
(*ICD-9*: 720.0. *ICD-10*: M45).

Other spondyloarthritides
(*ICD-9*: 720.1, 720.2, 720.8, and 720.9. *ICD-10*: M46.0, M46.1, M46.2, M46.3, M46.4, M46.5 M46.8, and M46.9).

Scleroderma
(*ICD-9*: 710.1. *ICD-10*: M34).

Sjogren's syndrome
(*ICD-9*: 710.2. *ICD-10*: M35.0).

Dermatomyositis
(*ICD-9*: 710.3. *ICD-10*: M33.1 and M33.9).

Polymyositis
(*ICD-9*: 710.4. *ICD-10*: M33.2).

Other connective tissue diseases
(*ICD-9*: 710.5, 710.8, and 710.9. *ICD-10*: M35.1, M35.2, M35.8, and M35.9).

Primary systemic vasculitis
(*ICD-9*: 446.0, 446.2, 446.4, 446.5, 446.7, and 447.6. *ICD-10*: D69.0, M31.0, M30.0, M31.3, M31.4, M31.5, M31.6, M31.7, M31.8, and M31.9).

Statistical analysis

The data were analyzed with SAS Version 8.3 software (SAS Institute Inc., Cary, NC, USA). Descriptive statistics were used to characterize the case and control cohorts. The sensitivity, specificity, and overall accuracy, positive, and negative predictive values of the seven case definitions for RA using administrative data were determined using the diagnosis made by rheumatology consultation as the gold standard. The extent of agreement was expressed by simple kappa coefficient (0.01–0.2 indicates slight agreement, 0.21–0.4 is fair, 0.41–0.6 is moderate, 0.61–0.8 is substantial agreement, and 0.81–0.99 is almost perfect agreement).[10]

Table 1 Demographic features of patients with RA and matched controls from health care administrative databases

Variable	RA cases	Controls	*P*-value
Number	535	2,140	
Age, mean (SD) (years)	56.0 (17.8)	56.0 (17.8)	NS*
Female sex (%)	67.9	67.9	NS*

Abbreviations: RA, rheumatoid arthritis; SD, standard deviation; NS, not significant.
Note: *, $P>0.05$.

Results

Cases and controls

From the administrative databases, a total of 535 RA case-admins were identified, all of whom had been seen by at least one rheumatologist in the Arthritis Center. In addition, 2,140 control-admins matched 4:1 to case-admins by age and sex were identified in the administrative databases that had an evaluation at the Arthritis Center (Table 1). Case-admins and control-admins were well matched for age and sex. These data provided the basis for validation of the case-admins' definitions for RA.

Identification of RA cases

The overall accuracy of the case definitions for the correct identification of RA case-admins (Table 2) varied between 68.9% and 82.9% with a kappa statistic between 0.26 and 0.53. The sensitivity and specificity of the case-admin definitions varied from 20.7% to 94.8% and 62.5% to 98.5%, respectively.

Impact of case definitions on incidence and prevalence rate of RA

To illustrate the potential impact of the different case-admin definitions, the estimated mean number of annual incident and prevalent cases of RA from 2002 to 2011 determined using the seven case-admin definitions are illustrated in Figure 1. The number of incident RA cases per year by case definition varied between 176 and 1,610. Given a mean population of 941,500 for Nova Scotia over the study period, this represents an annual incidence rate of 0.02%–0.17% for RA. The number of prevalent cases per year by case definition varied between 1,384 and 5,722 (0.15%–0.61% annual prevalent rate for Nova Scotia population).

Discussion

Administrative health care databases are repositories of clinical information, which may be used to evaluate the frequency of disease, utilization, and cost of health care resources in a population.[1] The value of any population health data set is influenced by a number of factors, such as the model of health care delivery, access to health services, and geographic stability of the population under study.[11–13] In addition, the validity of case definitions of disease and diagnostic groups is critical. Previous studies of RA have used a variety of methodological approaches to identify cases in administrative data sets, some of which have been validated through linkage to clinical data.[11] In the current study, we compared seven case definitions utilizing data from administrative health care databases in an urban/rural environment of ~1 million people, most of whom had access to universal health care. There were substantial differences between the case definitions in the accuracy of RA ascertainment and the consequent estimates of disease frequency. Our findings have implications for the use of administrative health care databases in the study of RA at a population health level.

Over the past 30 years, a number of studies have validated a variety of case definitions, consisting of single items or combinations thereof, for the identification of rheumatic diseases in administrative databases.[11,14–20] The results demonstrate a wide range in standard epidemiological measures, such as sensitivity, specificity, positive, and negative predictive values. The reason is due in part to not only the

Table 2 Agreement, sensitivity, specificity, accuracy, and predictive value of RA case definitions using administrative health care databases

Decision rule	Kappa (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)	Positive predicted value (95% CI)	Negative predicted value (95% CI)
#1 MacLean	0.52 (0.46, 0.56)	83.0 (79.8, 86.2)	80.8 (79.1, 82.5)	81.2 (79.4, 83.0)	51.9 (48.6, 55.3)	95.0 (94.0, 96.0)
#2 MacLean-Lacaille	0.42 (0.38, 0.46)	68.8 (64.9, 72.7)	80.8 (79.1, 82.5)	78.3 (76.5, 80.2)	47.2 (43.7, 50.7)	91.2 (90.0, 92.5)
#3 Shipton	0.53 (0.50, 0.57)	83.2 (80.0, 86.3)	81.6 (80.0, 83.2)	81.9 (80.2, 83.7)	53.0 (49.7, 56.4)	95.1 (94.1, 96.1)
#4 Hospitalization	0.26 (0.22, 0.31)	20.7 (17.3, 24.2)	98.5 (97.9, 99.0)	82.9 (91.5, 84.4)	77.1 (70.2, 83.9)	83.2 (83.2, 84.7)
#5 Rheumatologist	0.48 (0.44, 0.51)	87.7 (84.9, 90.5)	75.5 (73.7, 77.3)	77.9 (76.0, 80.0)	47.2 (44.1, 50.3)	96.1 (95.1, 97.0)
#6 Combination	0.49 (0.46, 0.52)	92.0 (89.7, 94.3)	74.3 (72.4, 76.1)	77.8 (75.8, 79.8)	47.2 (44.1, 50.2)	97.4 (96.6, 98.1)
#7 Single admin	0.37 (0.34, 0.40)	94.8 (92.9, 96.7)	62.5 (60.4, 64.5)	68.9 (66.5, 71.4)	38.7 (36.1, 41.3)	98.0 (97.2, 98.7)

Abbreviations: CI, confidence interval; RA, rheumatoid arthritis.

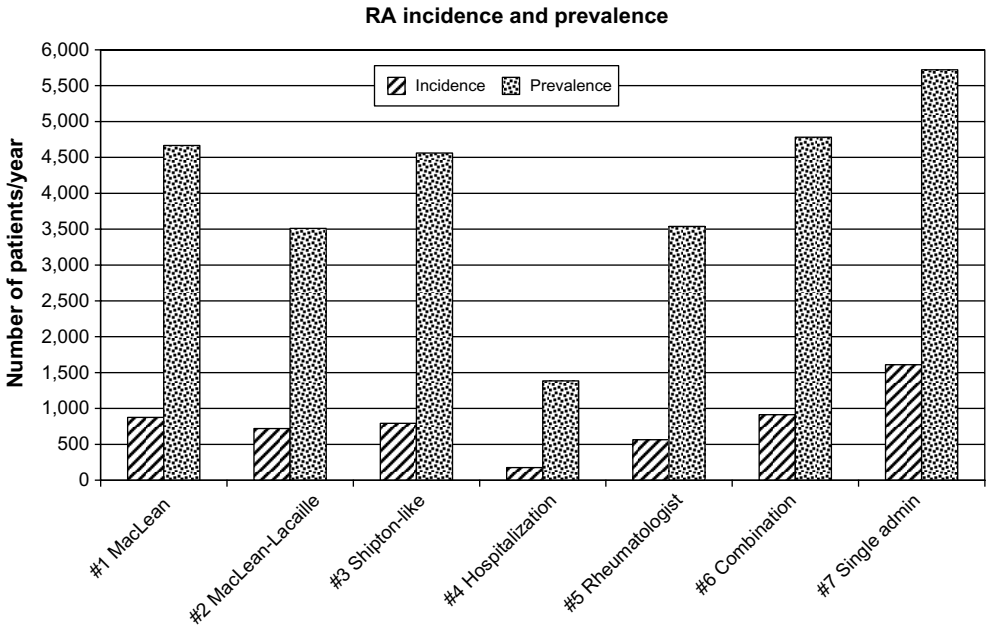

Figure 1 The mean annual incidence and prevalence rate of RA from 2002 to 2011 determined from health care administrative data using seven case definitions.
Abbreviation: RA, rheumatoid arthritis.

inherent differences between case definitions but also the methodological design of the validation strategy, including the selection of cases and controls (eg, general population or specialty clinics), diagnostic confirmation of cases (eg, patient self-report, classification criteria, or case definition), and blinding of assessors. In this study, cases and controls identified in the administrative databases were compared with the clinical diagnosis made by a rheumatologist.

A review of 12 studies[11] that validated case definitions for the identification of RA in administrative data sets indicates a wide range in sensitivity (8.3%–100%), specificity (27.0%–99.7%), positive (55.7%–95.0%), and negative (92.4%–100%) predictive values. In this study, variability in the performance of case definitions was also apparent, although less than that reported for RA in previous studies.

The prevalence rate of RA is felt to be between 0.50% and 1.0% of the general population,[2] and our findings (0.15%–0.61%) overlap with this albeit on the lower side. In a study from another large Canadian population, the prevalence rate of RA was 0.76%[7] compared to 0.37% in our population using the same case definition (#2 MacLean-Lacaille decision rule).[7] The difference is due in part to the exclusion of juvenile onset IA in our study but not in the study by Lacaille et al.[7] The lower disease prevalence rates for RA may also reflect the relative paucity of adult rheumatologists in Nova Scotia during the study period and reduced access to rheumatology services.

Differences in performance of the diagnostic case definitions have a significant impact on the estimation of the frequency of incident and prevalent RA cases. This is not altogether surprising. For example, given the nature of rheumatic diseases, it is to be expected that hospitalization data (case definition #4) would provide the lowest sensitivity for the detection of RA. On the other hand, the same case definition provides the highest specificity for RA cases. It is likely that this population includes patients with the greatest disease severity and the highest health care utilization and costs. Conversely, the case definition that relies solely on a single physician billing provides the highest sensitivity and lowest specificity for RA and likely includes patients with milder clinical disease phenotypes. Rather than a weakness, the variability in the performance of the different diagnostic case definitions is a potential strength, as it provides an opportunity to select a specific case definition a priori for its known sensitivity and specificity or to utilize a range of case definitions as a form of sensitivity analysis.

There are some limitations to the current study. First, although it would have been preferable not to restrict the validation of RA cases to patients assessed at a rheumatology clinic, the confirmation of cases and controls in primary care and nonrheumatology specialty settings was not feasible, in large part due to privacy legislation in Nova Scotia. Second, the criteria for selection of controls in our study preclude an assessment of the case definitions to distinguish between

RA and other IAs as the latter were deliberately excluded from the control group. Third some "false negatives" (ie, a patient with RA confirmed by a rheumatologist but not captured by any of the seven case definitions) may not have been included in the control groups, which, except for age and sex, were randomly matched 4:1 and excluded individuals with IA. However, the likelihood of a patient with RA not receiving a single physician billing diagnosis of RA over the 9 years of the study is low. This is one of the advantages of using seven case definitions for RA, which to our knowledge no other study has done. Fourth, the rheumatology clinic population is understandably enriched for RA cases, compared to the general population, which limits the reliability of estimating both positive and negative predictive values. These results were included to comply with the recent recommendations[11] arising from a systematic review of validation studies to identify rheumatic diseases in health administrative databases. Fifth, the generalizability of sensitivity and specificity in the study is limited to the use of RA case definitions based on the diagnostic codes from rheumatologists. Finally, it would be of interest to know the performance of these case definitions in the identification of other rheumatic disease entities.

Conclusion

The accuracy of case definitions for the identification of RA cases from rheumatology clinics using administrative health care databases is variable when compared to a rheumatologist's assessment. This has both advantages and disadvantages in the study of this chronic disease. The results of the current study provide a foundation to examine several important issues, including health care utilization and associated costs, in patients with RA.

Acknowledgments

Financial support for this study was provided by the John and Marian Quigley Endowment Fund for Rheumatology and by unrestricted funding from Glaxo Smith Kline, Canada and Roche Canada.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol*. 2005;58(4):323–337.
2. Gibofsky A. Epidemiology, pathophysiology, and diagnosis of rheumatoid arthritis: a synopsis. *Am J Manag Care*. 2014;20(7 Suppl): S128–S135.
3. Hallert E, Husberg M, Kalkan A, Skogh T, Bernfort L. Early rheumatoid arthritis 6 years after diagnosis is still associated with high direct costs and increasing loss of productivity: the Swedish TIRA project. *Scand J Rheumatol*. 2014;43(3):177–183.
4. 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 (Oxford)*. 2007; 46(12):1814–1818.
5. Ladouceur M, Rahme E, Pineau CA, Joseph L. Robustness of prevalence estimates derived from misclassified data from administrative databases. *Biometrics*. 2007;63(1):272–279.
6. Clarke AJ, Gulati P, Abraham SM. A cross-sectional audit of the uptake of seasonal and H1N1 influenza vaccination amongst patients with rheumatoid arthritis in a London hospital. *Clin Exp Rheumatol*. 2011;29(3):596.
7. Lacaille D, Anis AH, Guh DP, Esdaile JM. Gaps in care for rheumatoid arthritis: a population study. *Arthritis Rheum*. 2005;53(2): 241–248.
8. MacLean CH, Louie R, Leake B, et al. Quality of care for patients with rheumatoid arthritis. *JAMA*. 2000;284(8):984–992.
9. Shipton D, Glazier RH, Guan J, Badley EM. Effects of use of specialty services on disease-modifying antirheumatic drug use in the treatment of rheumatoid arthritis in an insured elderly population. *Med Care*. 2004;42(9):907–913.
10. Blackman NJ, Koval JJ. Interval estimation for Cohen's kappa as a measure of agreement. *Stat Med*. 2000;19(5):723–741.
11. Widdifield J, Labrecque J, Lix L, et al. Systematic review and critical appraisal of validation studies to identify rheumatic diseases in health administrative databases. *Arthritis Care Res (Hoboken)*. 2013; 65(9):1490–1503.
12. Goldfield N, Villani J. The use of administrative data as the first step in the continuous quality improvement process. *Am J Med Qual*. 1996; 11(1):S35–S38.
13. Schwartz RM, Gagnon DE, Muri JH, Zhao QR, Kellogg R. Administrative data for quality improvement. *Pediatrics*. 1999; 103(1 Suppl E):291–301.
14. Allebeck P, Ljungstrom K, Allander E. Rheumatoid arthritis in a medical information system: how valid is the diagnosis? *Scand J Soc Med*. 1983;11(1):27–32.
15. Gabriel SE. The sensitivity and specificity of computerized databases for the diagnosis of rheumatoid arthritis. *Arthritis Rheum*. 1994; 37(6):821–823.
16. Harrold LR, Yood RA, Andrade SE, et al. Evaluating the predictive value of osteoarthritis diagnoses in an administrative database. *Arthritis Rheum*. 2000;43(8):1881–1885.
17. 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):1594–1600.
18. Lin KJ, Garcia Rodriguez LA, Hernandez-Diaz S. Systematic review of peptic ulcer disease incidence rates: do studies without validation provide reliable estimates? *Pharmacoepidemiol Drug Saf*. 2011; 20(7):718–728.
19. Lix LM, Yogendran MS, Shaw SY, Burchill C, Metge C, Bond R. Population-based data sources for chronic disease surveillance. *Chronic Dis Can*. 2008;29(1):31–38.
20. Losina E, Barrett J, Baron JA, Katz JN. Accuracy of Medicare claims data for rheumatologic diagnoses in total hip replacement recipients. *J Clin Epidemiol*. 2003;56(6):515–519.