

BMJ Open Model-based methods for case definitions from administrative health data: application to rheumatoid arthritis

Kristine Kroeker,¹ Jessica Widdifield,^{2,3} Saman Muthukumarana,⁴ Depeng Jiang,¹ Lisa M Lix¹

To cite: Kroeker K, Widdifield J, Muthukumarana S, *et al.* Model-based methods for case definitions from administrative health data: application to rheumatoid arthritis. *BMJ Open* 2017;7:e016173. doi:10.1136/bmjopen-2017-016173

► Prepublication history for this paper is available online. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2017-016173>)

Received 29 January 2017

Revised 15 May 2017

Accepted 31 May 2017

ABSTRACT

Objective This research proposes a model-based method to facilitate the selection of disease case definitions from validation studies for administrative health data. The method is demonstrated for a rheumatoid arthritis (RA) validation study.

Study design and setting Data were from 148 definitions to ascertain cases of RA in hospital, physician and prescription medication administrative data. We considered: (A) separate univariate models for sensitivity and specificity, (B) univariate model for Youden's summary index and (C) bivariate (ie, joint) mixed-effects model for sensitivity and specificity. Model covariates included the number of diagnoses in physician, hospital and emergency department records, physician diagnosis observation time, duration of time between physician diagnoses and number of RA-related prescription medication records.

Results The most common case definition attributes were: 1+ hospital diagnosis (65%), 2+ physician diagnoses (43%), 1+ specialist physician diagnosis (51%) and 2+ years of physician diagnosis observation time (27%). Statistically significant improvements in sensitivity and/or specificity for separate univariate models were associated with (all p values <0.01): 2+ and 3+ physician diagnoses, unlimited physician diagnosis observation time, 1+ specialist physician diagnosis and 1+ RA-related prescription medication records (65+ years only). The bivariate model produced similar results. Youden's index was associated with these same case definition criteria, except for the length of the physician diagnosis observation time.

Conclusion A model-based method provides valuable empirical evidence to aid in selecting a definition(s) for ascertaining diagnosed disease cases from administrative health data. The choice between univariate and bivariate models depends on the goals of the validation study and number of case definitions.

INTRODUCTION

Administrative health data are widely used for research and surveillance studies because they are relatively inexpensive to access, cover entire populations and can be linked to create longitudinal patient-specific records of healthcare use. However, one limitation of administrative health data is their potentially low sensitivity and specificity for ascertaining

Strengths and limitations of this study

- Studies about the validity (ie, sensitivity and specificity) of disease case definitions for administrative health data typically rely on descriptive methods to select one or more case definitions for use. Our study proposes and demonstrates a model-based method that provides empirical evidence to support case definition development.
- Our method can be applied to diseases that are ascertained from diagnoses and/or prescription medication information in administrative health data, for which one or more validity measures are produced: sensitivity, specificity, positive predictive value, negative predictive value and summary measures such as Youden's index.
- A limitation of our method is that it cannot be applied to validation studies with a small number of case definitions (ie, <50 case definitions).

patients with chronic diseases.^{1–4} Therefore, validation studies are an essential tool for assessing data quality. A validation study compares cases ascertained from administrative health data with clinically confirmed cases and produces accuracy estimates (eg, sensitivity and specificity) for one or more case definitions.^{4,5} Many studies routinely test multiple case definitions. For example, in the rheumatoid arthritis (RA) validation literature, several studies reported more than 40 case definitions.^{6–8}

Selecting a single case definition from among the many that may be tested in a single study is not a straightforward process; sensitivity and specificity estimates often vary with case definition criteria such as the number of diagnosis codes, number of years of data used to ascertain cases and patient characteristics (eg, age and sex). Published guidelines recommend selecting a case definition by prioritising a single validity measure.⁹ Moreover, these guidelines recommend that validation studies report all case definitions and at least four validity measures.^{10 11} Thus,



CrossMark

¹Department of Community Health Sciences, University of Manitoba, Winnipeg, Canada

²Institute for Clinical Evaluative Sciences, Toronto, Canada

³Research Institute of the McGill University Health Centre, McGill University, Montreal, Canada

⁴Department of Statistics, University of Manitoba, Winnipeg, Canada

Correspondence to

Dr. Lisa M Lix; lisa.lix@med.umanitoba.ca

a single validation study may result in a large volume of case definition information. Many researchers rely solely on descriptive analyses to summarise these data and select a case definition from among those that are tested.^{1 3 6} However, the case definition with the highest diagnostic validity estimate may not be more accurate than the case definition with the next highest diagnostic validity estimate, due to sampling error in the estimates. Inferential methods can be used to support the selection of a case definition; they can provide valuable empirical evidence about the case definition criteria associated with validity estimates. However, to date, there have been no recommendations from published guidelines of the use of inferential methods to analyse diagnostic validity estimates.^{10 11}

We propose a model-based method to facilitate the selection of disease case definitions from validation studies for administrative health data. The objectives are to: (A) test administrative health data criteria associated with the validity of case definitions, and (B) compare competing models applied to case definition validity estimates. The model-based method is demonstrated for an RA validation study.

METHODS

Data source

Study data were from an RA validation study⁶ conducted using administrative data from 1 April 1991 to 31 March 2011 for patients from Ontario, Canada. Case definitions for administrative health data were developed using medical records for 450 patients from 18 rheumatology clinics as the gold standard. Physician billing claims, hospital discharge abstracts and emergency room (ER) records were used to develop case definitions for all patients; in addition, pharmacy data were used to develop case definitions for patients aged 65+ years.

The published study data reported on validity estimates for 61 case definitions. Validity estimates for an additional 87 case definitions not reported in the publication were provided by the first author. Thus, a total of 148 case definitions were available for analysis. Of this number, 57 case definitions (38.5%) were tested for individuals 20+ years and 91 case definitions (61.5%) were tested for individuals 65+ years. All case definitions tested in the 20+ years age group were also tested in the 65+ years age group. The remaining 34 case definitions for the 65+ years age group included prescription medication criteria.

Study variables

The case definitions were described using the following criteria (table 1): age group, number of diagnoses in hospital discharge records, number of diagnoses in ER records, number of diagnoses in physician records, number of specialist physician diagnoses, length of physician diagnosis observation time, ≥ 60 days of separation between physician diagnoses, exclusion criteria A, exclusion criteria B, number of RA-related medications

Table 1 RA case definition criteria and attributes

Case definition criteria	Attributes
Age group (years)	20+, 65+
Diagnoses in hospital discharge records	0, 1+
Diagnoses in ER records	0, 1+
Diagnoses in physician records	0, 1+, 2+, 3+
Specialist physician diagnoses	0, 1+
Length of physician diagnosis observation time	never, ≤ 1 year, ≤ 2 years, ≤ 5 years, unlimited
≥ 60 days between physician diagnoses	Yes, no
Exclusion criteria A	Yes, no
Exclusion criteria B	Yes, no
RA-related medications, including steroids	0, 1+
RA-related medications, excluding steroids	0, 1+

Exclusion criteria A excluded individuals with 2+ physician diagnosis codes with a different rheumatology diagnosis than an RA diagnosis; exclusion criteria B excluded individuals for whom the RA diagnosis was not confirmed by a specialist. Note: the attributes for age group, diagnoses in physician records and length of physician diagnosis observation time are not mutually exclusive. ER, emergency room; RA, rheumatoid arthritis.

including steroids and number of RA-related medications excluding steroids. Diagnoses in hospital discharge records and diagnoses in ER records were ascertained using an unlimited observation period. A case definition with no physician diagnoses corresponds with having no physician diagnosis observation time. The exclusion criteria A was defined by the authors as follows: exclude individuals with at least two physician diagnosis codes with a different rheumatology diagnosis to an RA diagnosis; this includes osteoarthritis, gout, polymyalgia rheumatic, other seronegative spondyloarthropathy, ankylosing spondylitis, psoriasis, synovitis/tenosynovitis/bursitis, connective tissue disorder, vasculitis and others. The exclusion criteria B was defined by the authors as follows: exclude individuals who had an RA diagnosis code not confirmed by a specialist. The RA-related medication criteria were set to missing for the case definitions applied to the 20+ age group, because medication data were not available for this age group.

The study dataset included each case definition as an observation. The case definition criteria and estimates of sensitivity and specificity were included as variables. Youden's index (ie, sensitivity + specificity - 1)¹² was calculated from the estimates of sensitivity and specificity.

Statistical analyses

Descriptive analyses of the case definition attributes and the estimates of sensitivity, specificity and Youden's index were conducted using frequencies, percentages and means to inform the model fitting process. All criteria

were treated as ordinal measures. Spearman correlation coefficients were used to identify potential collinearity (defined as a correlation of 0.70 or greater¹³) among the case definition criteria.

The following models were fit to the data based on previous research¹⁴: (A) separate univariate fixed-effects models for sensitivity and specificity, (B) univariate fixed-effects model for Youden's index and (C) bivariate (ie, joint) mixed-effects model for sensitivity and specificity. For the univariate models, sensitivity, specificity or Youden's index were the outcome variables, and the case definition criteria were covariates. The bivariate mixed-effects model jointly modelled sensitivity and specificity as the outcome variables, and the case definition criteria were covariates. In the bivariate model, estimates of sensitivity and specificity were treated as repeated measures to account for their dependence.

Each covariate (ie, case definition criteria) was first tested in unadjusted models. Multivariable models were subsequently fit to the data; only the covariates that were statistically significant at $\alpha=0.01$ and explained more than 1% of the variation in the unadjusted models (based on the pseudo R^2 statistic¹⁵) were retained. The pseudo R^2 statistic was calculated using the likelihood statistics from the unadjusted model and the null model (ie, model with no covariates) using McFadden's method.¹⁵ A nominal $\alpha=0.01$, based on the Bonferroni correction, was used to evaluate statistical significance in the multivariable model to limit the overall probability of a Type I error.¹⁶ The adjusted models were fit to the data for all age groups ($n=148$) and then fit to the data only for the 65+ age group ($n=91$). Univariate models were reported with the percent of explained variance. All model estimates were reported with 99% CIs.

The data were modelled using a beta distribution with a logit link function as recommended in previous research.¹⁴ The mixed-effects bivariate model used the Cholesky decomposition to ensure that the estimated variance-covariance matrix of the random effects was positive semidefinite and the model converged.¹⁷ All analyses were conducted using SAS V.9.3.¹⁸

RESULTS

Descriptive analyses

As shown in table 2, two-thirds (64.9%) of the case definitions that were applied to the data for all age groups used 1+ hospital discharge record in an unlimited observation period as a criterion. At least one diagnosis in ER records in an unlimited observation period was a criterion of 6.8% of the case definitions. Physician claims were used in 94.6% of the case definitions. The case definitions used physician diagnosis observation periods of never (ie, when physician claims were not used, 5.4%), ≤ 1 year (25.7%), ≤ 2 years (27.0%), ≤ 5 years (29.7%) and an unlimited physician diagnosis observation period (12.2%) to ascertain physician diagnoses. At least one specialist diagnosis was included as a criterion

in half of the case definitions (51.4%). A time separation of ≥ 60 days between two physician diagnoses was used in 14 (9.5%) case definitions. Exclusion criteria A and B were infrequently used in the case definitions (6.8% and 5.4%, respectively). Of the 91 case definitions for the 65+ age group, 11.5% required 1+ RA-related medication including steroids to ascertain cases and 11.5% of case definitions used 1+ RA-related medication excluding steroids to ascertain cases. Compared with the case definitions for the 65+ years age group, the case definitions for the 20+ age group had slightly lower average estimates of sensitivity (20+ years: 90.9 and 65+ years: 91.3), specificity (20+ years: 82.2 and 65+ years: 86.1) and Youden's index (20+ years: 73.1 and 65+ years: 77.4).

The following case definition criteria were highly correlated (data not shown): exclusion criteria A and B ($r=0.89$; $p<0.0001$), exclusion criteria A and ≥ 60 days of separation between physician claims ($r=0.83$; $p<0.0001$) and exclusion criteria B and ≥ 60 days of separation between physician claims ($r=0.74$; $p<0.0001$). These combinations of case definition criteria were not included in the same model; rather, one criterion from each pair was used in a model at a time.

Inferential analyses

Table 3 reports the percent pseudo R^2 statistics for the univariate and bivariate unadjusted models for all case definitions. The number of physician diagnoses and physician diagnosis observation time were the two criteria that explained the most variation in all models (19.2%–82.6% and 16.0%–77.4%, respectively). In the unadjusted univariate models, sensitivity was significantly associated ($p<0.01$) with the number of physician diagnoses, length of physician diagnosis observation time, ≥ 60 days of separation between physician diagnoses, exclusion criteria A and B and number of RA-related medications excluding steroids. Specificity was significantly associated ($p<0.01$) with the number of physician diagnoses, number of specialist diagnoses, length of physician diagnosis observation time, ≥ 60 days of separation between physician diagnoses and number of RA-related medications excluding steroids. The unadjusted univariate models also revealed that Youden's index was significantly associated ($p<0.01$) with the number of diagnoses in ER records, number of physician diagnoses, number of specialist diagnoses, length of physician diagnosis observation time, ≥ 60 days of separation between physician diagnoses and exclusion criteria A and B. The joint estimates of sensitivity and specificity were significantly associated ($p<0.01$) with the number of hospital diagnoses, number of physician diagnoses, number of specialist diagnoses, length of physician diagnosis observation time, ≥ 60 days of separation between physician diagnoses, exclusion criteria A and B and number of RA-related medications excluding steroids in the bivariate models.

When all case definitions ($n=148$) were considered, the adjusted univariate model of sensitivity showed that increasing the length of physician diagnosis observation

Table 2 Descriptive statistics for RA case definition criteria (n=148)

Case definition criteria	Frequency (%)	Sensitivity Mean (SD)	Specificity Mean (SD)	Youden's index Mean (SD)
Cohort				
20+ years	57 (38.5)	90.9 (17.1)	82.2 (7.5)	73.1 (15.3)
65+ years	91 (61.5)	91.3 (15.2)	86.1 (7.4)	77.4 (13.9)
Hospital diagnoses				
0	52 (35.1)	96.7 (2.4)	83.7 (6.8)	80.4 (5.9)
1+	96 (64.9)	88.1 (19.0)	85.1 (8.0)	73.2 (17.0)
ER diagnoses				
0	138 (93.2)	92.6 (12.8)	84.8 (6.8)	77.4 (12.4)
1+	10 (6.8)	71.4 (34.1)	81.3 (14.9)	52.7 (22.1)
Physician diagnoses				
0	8 (5.4)	30.7 (4.4)	96.9 (1.1)	27.7 (5.0)
1+	16 (10.8)	98.4 (2.4)	71.3 (9.7)	69.7 (8.7)
2+	64 (43.2)	93.5 (9.2)	83.2 (4.8)	76.7 (10.6)
3+	60 (40.5)	94.7 (2.1)	88.0 (3.7)	82.8 (3.0)
Specialist diagnoses				
0	72 (48.7)	85.8 (21.4)	81.8 (9.5)	67.5 (17.2)
1+	76 (51.4)	96.2 (2.7)	87.3 (3.8)	83.5 (2.8)
Physician diagnosis observation time				
Never	8 (5.4)	30.7 (4.4)	96.9 (1.1)	27.7 (5.0)
≤1 year	38 (25.7)	94.9 (2.8)	87.5 (5.4)	82.4 (3.9)
≤2 years	40 (27.0)	94.6 (7.3)	85.0 (4.4)	79.6 (8.5)
≤5 years	44 (29.7)	93.2 (8.7)	84.1 (4.5)	77.3 (10.6)
Unlimited	18 (12.2)	97.6 (2.9)	73.1 (10.6)	70.7 (8.8)
≥60 days between physician diagnoses				
Yes	14 (9.5)	79.8 (11.1)	80.5 (2.9)	60.3 (10.8)
No	134 (90.5)	92.3 (15.8)	85.0 (7.8)	77.3 (13.9)
Exclusion criteria A				
Yes	10 (6.8)	73.2 (2.6)	80.8 (1.3)	54.0 (2.6)
No	138 (93.2)	92.4 (15.6)	84.9 (7.8)	77.3 (13.7)
Exclusion criteria B				
Yes	8 (5.4)	73.7 (2.6)	80.5 (1.2)	54.2 (2.8)
No	140 (94.6)	92.1 (15.7)	84.8 (7.8)	77.0 (14.0)
RA-related medications, including steroids				
0	74 (50.0)	91.3 (15.2)	86.5 (7.1)	77.8 (14.0)
1+	17 (11.5)	91.1 (15.5)	84.5 (8.4)	75.8 (13.5)
Missing	57 (38.5)	90.9 (17.1)	82.2 (7.5)	73.1 (15.3)
RA-related medications, excluding steroids				
0	74 (50.0)	91.9 (15.3)	84.9 (7.5)	76.8 (14.0)
1+	17 (11.5)	88.7 (14.8)	91.2 (3.7)	79.9 (13.5)
Missing	57 (38.5)	90.9 (17.1)	82.2 (7.5)	73.1 (15.3)

Exclusion criteria A excluded individuals with 2+ physician diagnosis codes with a different rheumatology diagnosis than an RA diagnosis; exclusion criteria B excluded individuals whose RA diagnosis was not confirmed by a specialist. Note: the attributes for age group, diagnoses in physician records and length of physician diagnosis observation time are not mutually exclusive. ER, emergency room; RA, rheumatoid arthritis.

Table 3 Percent pseudo- R^2 statistics for unadjusted univariate models of sensitivity, specificity, univariate model of Youden's index and bivariate model of sensitivity and specificity (n=148)

Case definition criteria	Univariate model sensitivity	Univariate model specificity	Univariate model Youden's index	Bivariate model
Physician diagnoses	28.8*	36.9*	82.6*	19.2*
Physician diagnosis observation time	25.2*	28.9*	77.4*	16.0*
Specialist diagnoses	0.6	2.1*	25.5*	8.0*
≥60 days between physician diagnoses	1.4*	1.9*	9.2*	7.7*
Exclusion criteria A	4.1*	1.3	12.7*	9.5*
Exclusion criteria B	3.1*	1.1	9.7*	7.0*
Hospital diagnoses	1.0	0.7	3.1	1.3*
ER diagnoses	0.2	0.02	13.9*	1.3
RA-related medications, excluding steroids	2.4*	4.5*	0.6	1.9
RA-related medications, including steroids	1.2	0.2	0.4	0.8

Exclusion criteria A excluded individuals with 2+ physician diagnosis codes with a different rheumatology diagnosis than an RA diagnosis; exclusion criteria B excluded individuals whose RA diagnosis was not confirmed by a specialist.

*Indicate estimates that were statistically significant at $\alpha=0.01$.

ER, emergency room; RA, rheumatoid arthritis.

time to unlimited compared from 1 year was associated with a statistically significant increase in sensitivity, and ≥60 days of separation between physician diagnoses significantly decreased sensitivity (figure 1). Similar relationships were found for the models for the case definitions from the 65+ years, where prescription medication data were available (n=91).

When all case definitions (n=148) were considered, the univariate model of specificity showed that using 2+ and 3+ physician diagnoses were associated with a statistically significant increase in specificity compared with one physician diagnosis, increasing the physician diagnosis observation time to ≤2 years and ≤5 years from 1 year significantly decreased specificity and 1+ specialist diagnosis significantly increased specificity (figure 1). When only the case definitions for the 65+ age group (n=91) were considered, the results showed similar relationships. Also, the number of RA-related medications excluding steroids was associated with a statistically significant increase in specificity.

Based on the unadjusted models, the univariate model with all case definitions (n=148) for Youden's index included the case definition criteria of number of diagnoses in ER records, number of physician diagnoses, physician diagnosis observation time, number of specialist visits and ≥60 days of separation between physician diagnoses. However, the number of ER records and physician diagnosis observation time criteria were not statistically significant and model fit improved when they were removed (data not shown). The adjusted univariate model of Youden's index showed that using 2+ and 3+ physician diagnoses to ascertain cases significantly increased Youden's index compared with 1+ physician diagnosis, 1+ specialist diagnosis significantly increased Youden's index and a time separation ≥60 days between diagnoses significantly decreased Youden's index (figure 2). When

the case definitions for the 65+ population (n=91) were considered, a similar pattern emerged. Using the number of RA-related medications excluding steroids to ascertain RA cases resulted in a statistically significant increase in Youden's index.

When all case definitions were analysed using the adjusted bivariate model, 2+ and 3+ physician diagnoses were associated with a statistically significant increase in specificity and no association with sensitivity compared with 1+ physician diagnosis (figure 3). Increasing the number of physician diagnosis observation years from 1 year to ≤2 years, ≤5 years and unlimited observation period were associated with a statistically significant increase in sensitivity. Increasing the number of physician diagnosis observation years from 1 year to ≤2 years and ≤5 years were associated with a statistically significant decrease in specificity. Using ≥60 days of separation between physician diagnoses was associated with a statistically significant decrease in sensitivity but no significant change on specificity. Increasing the number of specialist diagnoses significantly increased specificity. When only the case definitions applied to the 65+ age group (n=91) were analysed, the relationships in the all age groups model remained statistically significant. Including 1+ RA-related medications excluding steroids decreased sensitivity and increased specificity.

DISCUSSION

This study applied regression models in a secondary analysis of administrative health data to identify case definition criteria associated with validity estimates from a study about RA case definitions. Based on the results of the adjusted univariate model, sensitivity was associated with the number of physician diagnoses, physician diagnosis observation time and length of time between physician

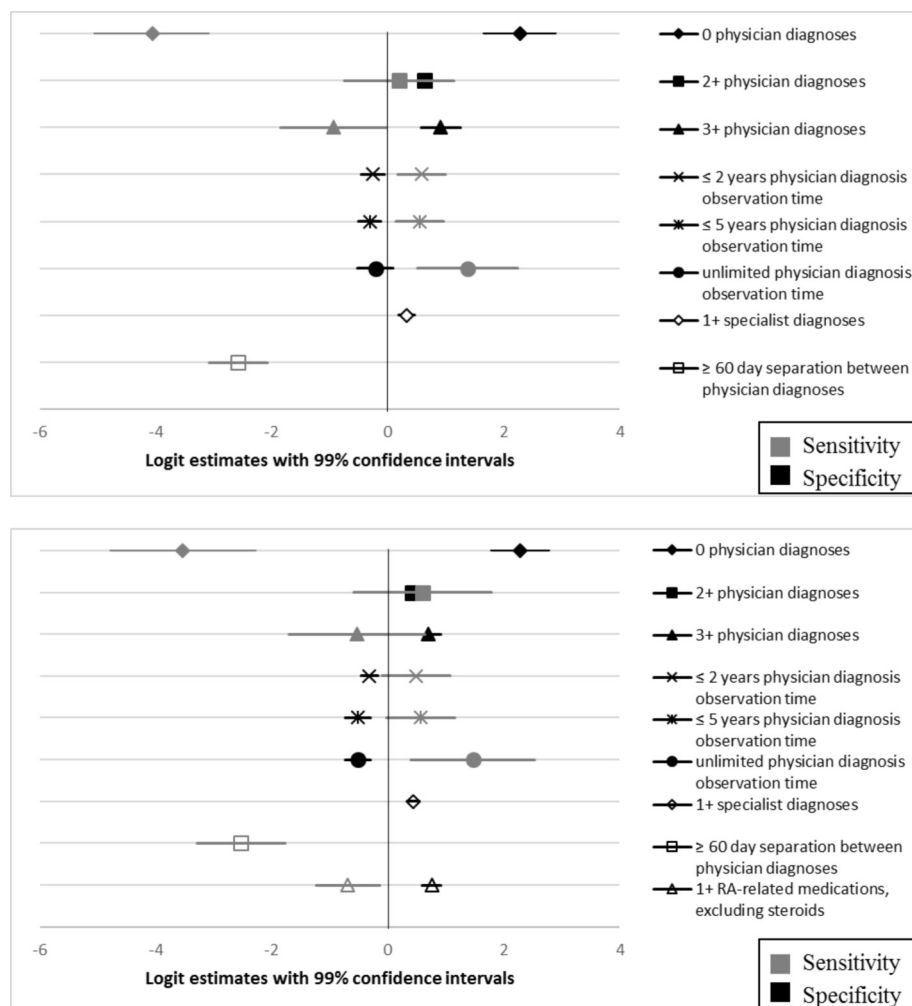


Figure 1 Logit estimates and 99% CIs for adjusted univariate models of sensitivity and specificity for case definitions applied to: (A) all age groups and (B) 65+ age group. Reference categories were: 1+ physician diagnoses, ≤ 1 year physician diagnosis observation time, 0 specialist diagnoses, no 60+ day separation between physician diagnoses and 0 RA-related medications excluding steroids. RA, rheumatoid arthritis.

diagnoses. Based on the results of the univariate model, specificity was associated with the number of physician diagnoses, physician diagnosis observation time, specialist diagnoses and RA-related medications excluding steroids. Based on the univariate model results, Youden's index was associated with the number of physician diagnoses, specialist diagnoses, length of time between physician diagnoses and number of RA-related medications excluding steroids. For the bivariate model, sensitivity was associated with the number of physician diagnoses, physician diagnosis observation time, amount of time between physician diagnoses and number of RA-related medications excluding steroids. In this same model, specificity was associated with the number of physician diagnoses, physician diagnosis observation time, specialist diagnoses and RA-related medications excluding steroids.

All of the models resulted in similar performance in our numeric example, but this may not always be the case. Selection of one model over competing alternatives depends on the study goals and the number of case definitions. Overall, however, the bivariate model

is recommended when the number of case definitions is large and sensitivity and specificity are moderately or highly correlated. The univariate model applied to Youden's index is recommended when the researcher places equal weight on maximising sensitivity and specificity.^{12 19–21} However, Youden's index can result in the same estimate for different combinations of sensitivity and specificity. Thus, univariate models applied separately to sensitivity and specificity are recommended when the researcher does not place equal weight on these validity measures.²²

A validation study is typically used to produce recommendations about selecting one or more case definitions for maximum accuracy in ascertaining disease cases. Using our model-based method, one can identify the case definition criteria associated with one or more measures of validity and use them to construct a case definition. Based on the univariate models for all age groups, the recommended RA case definition has the following attributes: (A) 2+ physician diagnoses, (B) unlimited physician diagnosis observation time and (C) 1+ specialist diagnosis. At

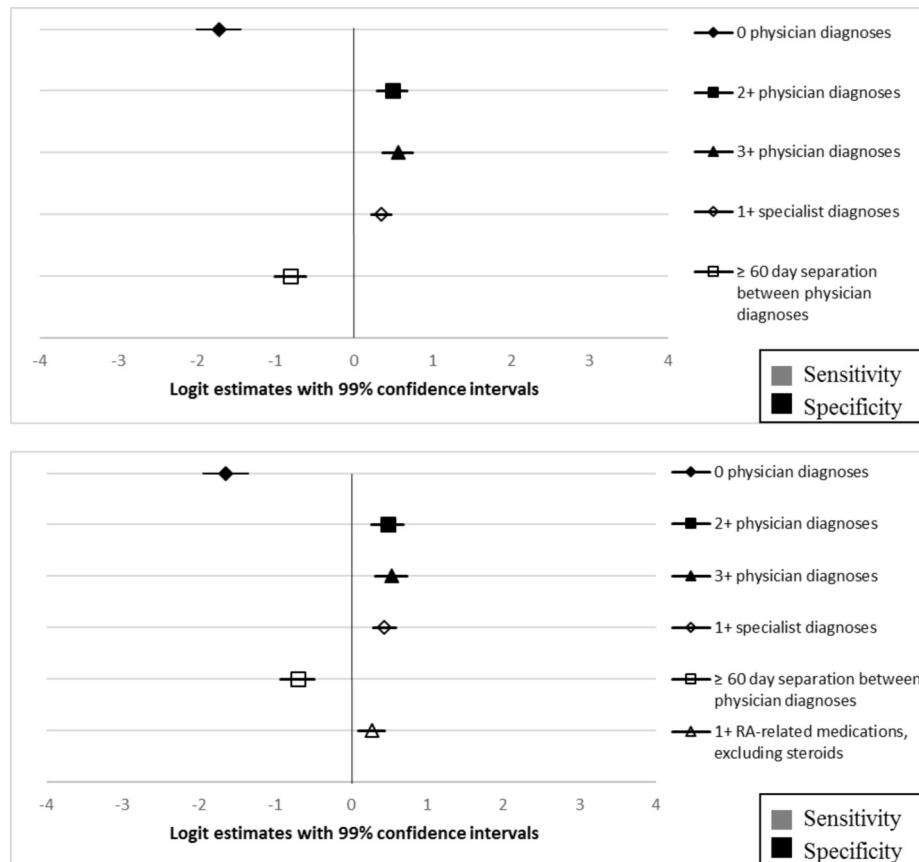


Figure 2 Logit estimates and 99% CIs for adjusted univariate models of Youden's index for case definitions applied to: (A) all age groups and (B) 65+ age group. Reference categories were: 1+ physician diagnoses, 0 specialist diagnoses, no 60+ day separation between physician diagnoses and 0 RA-related medications excluding steroids. RA, rheumatoid arthritis.

least two physician diagnoses and 1+ specialist diagnosis were associated with specificity but not with sensitivity. An unlimited observation time was significantly associated with improvements in sensitivity. For the 65+ age group, the recommended RA case definition had the same case definition attributes and also included 1+ RA-related medication excluding steroids. The univariate model of Youden's index and bivariate model of sensitivity and specificity produced similar results.

The recommended case definition based on the univariate sensitivity and specificity models is simpler than the recommended case definition from Widdifield *et al.*,⁶ however, both case definitions produce similar diagnostic accuracy estimates. The primary difference between the two recommended case definitions is that Widdifield *et al* recommended using one diagnosis in hospital discharge records to ascertain cases, while our model-based approach did not support this. Our recommendation derived from a model-based approach might lead to subsequent reanalysis of the original validation data to produce estimates of sensitivity and specificity for the model-supported case definition.

The case definition with the highest sensitivity or specificity estimates may not be significantly more accurate than other case definitions. A model-based approach provides empirical evidence about the case definition

criteria that are associated with significant increases/decreases in validity estimates.

While this research focused on inferential techniques for diagnostic validation studies, design of such studies is also an important consideration to ensure that the effects of the criteria can be accurately estimated. Ensuring that all possible combinations of criteria are investigated is an important consideration.²³ When a criterion is included in only a small number of case definitions, the power to detect the effect of the criterion on diagnostic validity estimates may be low.

This study has some limitations. Other parametric and non-parametric models have been proposed to combine estimates from a single case definition across multiple studies, such as copulas techniques,^{14 24} mixture models^{25 26} and mixed models of summary receiver operating characteristic curves.^{27 28} However, the models selected for this study have many applications and are most likely to be familiar to researchers. A beta distribution may not always be an appropriate choice for Youden's index, because this index can, in theory, range from -1 to +1. However, in practice, values of Youden's index less than zero are rare.

Inferential methods cannot be applied to the estimates from all diagnostic validation studies; overfitting the data may be a problem when the number of case definitions is small, or when the number of case definition criteria is

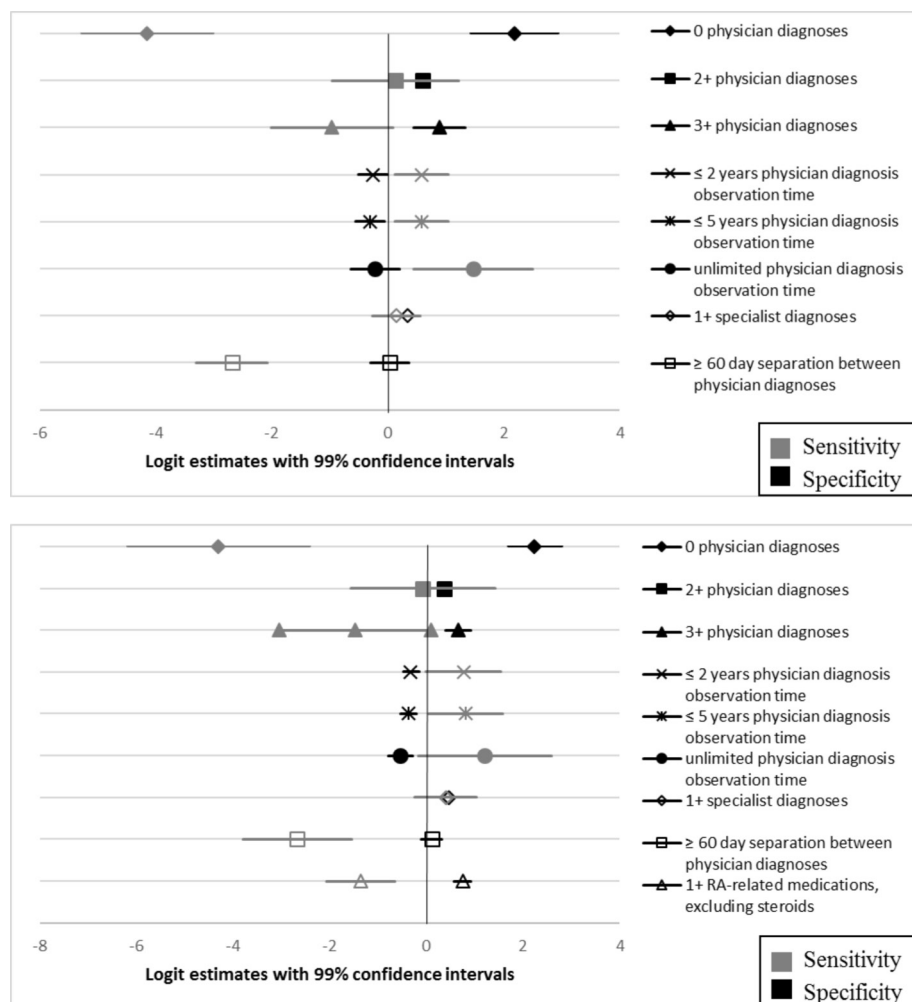


Figure 3 Logit estimates and 99% CIs for adjusted bivariate models of sensitivity and specificity for case definitions applied to: (A) all age groups and (B) 65+ age group. Reference categories were: 1+ physician diagnoses, ≤ 1 year physician diagnosis observation time, 0 specialist diagnoses, no 60+ day separation between physician diagnoses and 0 RA-related medications excluding steroids. RA, rheumatoid arthritis.

large relative to the total number of definitions. Green²⁹ suggested a minimum of 50 observations plus eight observations for every parameter estimated from a multiple regression model. Based on this, our model-based approach would require a minimum of 50 case definitions, and preferably more, in order to be implemented.²⁹ In such cases, descriptive analyses must be relied on to select a case definition. Lastly, the validation study design may limit the ability to test interactions between criteria. Testing interaction effects would require as many combinations as possible of the criteria to be included in the validation study to allow for reasonable model power.

This study has a number of strengths. The models applied to the case definition data from a single study have previously been used in meta-analyses to combine diagnostic validity estimates for a single case definition across multiple studies.^{14 30 31} Second, these methods enable modelling of case definitions from published validation studies and when the individual-level administrative health data are not available. Another strength of this study was that the effects of publication bias on the

study results was minimised by analysing all case definition data provided by the study authors. Finally, the methods used in this study can be applied to other chronic diseases and other diagnostic validity measures such as positive predictive value and/or negative predictive value.

CONCLUSION

This study applied model-based methods to a single validation study to select case definition criteria associated with validity measures such as sensitivity and specificity. The model-based results can be used by researchers to empirically guide the selection of a case definition for implementation in subsequent cohort studies and surveillance initiatives.^{32 33} Empirical methods can be used to quantify the magnitude of change in estimates of accuracy associated with different case definition criteria. This research contributes to accurate methods for using administrative health data to study chronic diseases such as RA.

Correction notice This paper has been amended since it was published Online

First. Owing to a scripting error, some of the publisher names in the references were replaced with 'BMJ Publishing Group'. This only affected the full text version, not the PDF. We have since corrected these errors and the correct publishers have been inserted into the references.

Contributors KK led the conception and design of the study, analysis and interpretation of the data and drafted the article. LML provided guidance on the conception and design of the study, assisted in the analysis and interpretation of the data and was involved in revising the article. JW provided access to the study data, assisted in interpretation of the data and was involved in revising the article. DJ and SM provided guidance on the conception and design of the study, assisted in the analysis and interpretation of the data and were involved in revising the article. All authors read and approved the final manuscript.

Funding The first author was supported by the Canadian Institutes of Health Research (CIHR) Drug Safety and Effectiveness Network grant TD3.137716 through the scholarship from the Canadian Network for Advanced Interdisciplinary Methods for comparative effectiveness research (CAN.AIM) team. This work was supported by the Canadian Institutes of Health Research (CIHR) (www.cihr.irsc.gc.ca) through Canadian Master's Scholarships funding from CIHR to the first author.

Competing interests None declared.

Patient consent This study does not involve human subjects.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The datasets used and analysed during the current study are available from JW on reasonable request.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2017. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

1. Quan H, Khan N, Hemmelgarn BR, *et al.* Validation of a case definition to define hypertension using administrative data. *Hypertension* 2009;54:1423–8.
2. Hux JE, Ivis F, Flintoft V, *et al.* Diabetes in Ontario: determination of prevalence and incidence using a validated administrative data algorithm. *Diabetes Care* 2002;25:512–6.
3. Tu K, Campbell NR, Chen ZL, *et al.* Accuracy of administrative databases in identifying patients with hypertension. *Open Med* 2007;1:e18–26.
4. 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;61:1250–60.
5. Lix LM, Yan L, Blackburn D, *et al.* Validity of the RAI-MDS for ascertaining diabetes and comorbid conditions in long-term care facility residents. *BMC Health Serv Res* 2014;14:17.
6. Widdifield J, Bernatsky S, Paterson JM, *et al.* Accuracy of canadian health administrative databases in identifying patients with rheumatoid arthritis: a validation study using the medical records of rheumatologists. *Arthritis Care Res* 2013;65:n/a–91.
7. Widdifield J, Bombardier C, Bernatsky S, *et al.* An administrative data validation study of the accuracy of algorithms for identifying rheumatoid arthritis: the influence of the reference standard on algorithm performance. *BMC Musculoskelet Disord* 2014;15:216.
8. Benchimol EI, Guttman A, Mack DR, *et al.* Validation of international algorithms to identify adults with inflammatory bowel disease in health administrative data from Ontario, Canada. *J Clin Epidemiol* 2014;67:887–96.
9. Chubak J, Pocobelli G, Weiss NS. Tradeoffs between accuracy measures for electronic health care data algorithms. *J Clin Epidemiol* 2012;65:343–9.
10. Benchimol EI, Manuel DG, To T, *et al.* Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. *J Clin Epidemiol* 2011;64:821–9.
11. Whiting P, Rutjes AW, Reitsma JB, *et al.* The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003;3:25.
12. Youden WJ. Index for rating diagnostic tests. *Cancer* 1950;3:32–5.
13. Tabachnick BG, Fidell LS. *Using multivariate statistics*. 2nd ed. New York: Harper & Row, 1989:p. 87–8.
14. Kuss O, Hoyer A, Solms A. Meta-analysis for diagnostic accuracy studies: a new statistical model using beta-binomial distributions and bivariate copulas. *Stat Med* 2014;33:17–30.
15. Domencich TA, McFadden D. *Urban travel demand: a behavioral analysis*. Amsterdam: North-Holland Publishing Company, 1975.
16. Shaffer JP. Multiple hypothesis testing. *Annu Rev Psychol* 1995;46:561–84.
17. Menke J. Bivariate random-effects meta-analysis of sensitivity and specificity with SAS PROC GLIMMIX. *Methods Inf Med* 2010;49:54–64.
18. SAS Institute Inc. *SAS/STAT 9.3 user's Guide*. Cary, NC, 2011.
19. Marrie RA, Yu BN, Leung S, *et al.* Rising prevalence of vascular comorbidities in multiple sclerosis: validation of administrative definitions for diabetes, hypertension, and hyperlipidemia. *Mult Scler* 2012;18:1310–9.
20. Butt DA, Tu K, Young J, *et al.* A validation study of administrative data algorithms to identify patients with parkinsonism with prevalence and incidence trends. *Neuroepidemiology* 2014;43:28–37.
21. Tu K, Wang M, Jaakkimainen RL, *et al.* Assessing the validity of using administrative data to identify patients with epilepsy. *Epilepsia* 2014;55:335–43.
22. Zaslavsky AM, Shaul JA, Zaboriski LB, *et al.* Combining health plan performance indicators into simpler composite measures. *Health Care Financ Rev* 2002;23:101–15.
23. Vanniyasingam T, Cunningham CE, Foster G, *et al.* Simulation study to determine the impact of different design features on design efficiency in discrete choice experiments. *BMJ Open* 2016;6:e011985.
24. Hoyer A, Kuss O. Meta-analysis of diagnostic tests accounting for disease prevalence: a new model using trivariate copulas. *Stat Med* 2015;34:1912–24.
25. Schlattmann P, Verba M, Dewey M, *et al.* Mixture models in diagnostic meta-analyses—clustering summary receiver operating characteristic curves accounted for heterogeneity and correlation. *J Clin Epidemiol* 2015;68:61–72.
26. Böhning D, Böhning W, Holling H. Revisiting Youden's index as a useful measure of the misclassification error in meta-analysis of diagnostic studies. *Stat Methods Med Res* 2008;17:543–54.
27. Doebler P, Holling H, Böhning D. A mixed model approach to meta-analysis of diagnostic studies with binary test outcome. *Psychol Methods* 2012;17:418–36.
28. Hamza TH, Reitsma JB, Stijnen T. Meta-analysis of diagnostic studies: a comparison of random intercept, normal-normal, and binomial-normal bivariate summary ROC approaches. *Med Decis Making* 2008;28:639–49.
29. Green SB. How many Subjects does It take to do A regression analysis. *Multivariate Behav Res* 1991;26:499–510.
30. Riley RD, Abrams KR, Sutton AJ, *et al.* Bivariate random-effects meta-analysis and the estimation of between-study correlation. *BMC Med Res Methodol* 2007;7:3.
31. Reitsma JB, Glas AS, Rutjes AW, *et al.* Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;58:982–90.
32. Widdifield J, Paterson JM, Bernatsky S, *et al.* The epidemiology of rheumatoid arthritis in Ontario, Canada. *Arthritis Rheumatol* 2014;66:786–93.
33. Widdifield J, Moura CS, Wang Y, *et al.* The Longterm effect of early intensive treatment of seniors with rheumatoid Arthritis: a comparison of 2 Population-based Cohort studies on Time to Joint Replacement Surgery. *J Rheumatol* 2016;43:861–8.