

Development of an indicator of smoking status for people with multiple sclerosis in administrative data

Ruth Ann Marrie, MD, PhD , Qier Tan, PhD, Okechukwu Ekuma, MSc and James J Marriott, MD, MSc*Multiple Sclerosis Journal—
Experimental, Translational
and Clinical*

January–March 2022, 1–5

DOI: 10.1177/
20552173221074296© The Author(s), 2022.
Article reuse guidelines:
sagepub.com/journals-
permissions

Abstract

Background: Administrative data lack health behavior information.**Methods:** We developed an administrative case definition for past or current ('ever smoking') in 1320 individuals with MS from Manitoba, Canada. Candidate indicators for 'ever smoked' included smoking cessation medications, and diagnosis codes for tobacco use and chronic obstructive pulmonary disease, using variable lookback periods.**Results:** When compared to self-reported smoking status, the case definition incorporating all indicators over a lifetime lookback period had a sensitivity of 31.98%, and positive predictive value of 78.26%.**Conclusion:** This smoking status definition could only partially control for confounding due to smoking because of the low sensitivity.**Keywords:** Administrative data, smoking, multiple sclerosis

Date received: 11 December 2021; accepted: 3 January 2022

Introduction

Administrative (health) claims data are frequently used to study epidemiology and health services use in multiple sclerosis (MS). Although administrative data have several strengths, they usually lack information about health behaviors such as smoking. Prior studies that have tested the validity of claims-based determination of smoking status using American data sources have suggested limited sensitivity but high specificity.^{1,2} However, variations in health system structure, billing and diagnostic coding practices mean that case definitions from one health system may perform differently in another system. Performance of case definitions may also vary from one clinical population to another. We tested an administrative case definition for 'ever smoked' status in an MS population, in Manitoba, Canada, a region with a universal, publicly funded health system.

Methods

As described elsewhere,³ we used administrative databases held in the Manitoba Population Data Repository at the Manitoba Centre for Health Policy. Databases used included the Population

Registry (date of birth, sex, health care coverage, postal code of residence); Discharge Abstract Database (hospitalizations including dates and diagnoses recorded using International Classification of Disease [ICD]-9th edition-clinical modification/10th edition-Canadian adaptation); Medical Services (physician visits, ICD-9-CM coded diagnosis); and Drug Program Information Network (DPIN, community-dispensed prescriptions including name, drug identification number (DIN), dispensation date). Linkage of postal code to census data provided area-level socioeconomic status based on annual household income.

We applied a validated case definition to identify Manitobans with MS from April 1, 1984–March 31, 2017. This definition required ≥ 3 health care contacts (hospitalizations, physician visits, prescription claims in any combination) and has a high sensitivity (99.5%), specificity (98.5%) and positive predictive value (PPV, 99.5%). The date of the first demyelinating disease claim (e.g. optic neuritis) was designated as the index (diagnosis) date. These data were linked to linked to smoking status information (dichotomized as ever versus never) obtained via questionnaire during clinic visits from consenting

Correspondence to:

Ruth Ann Marrie,
Department of Internal
Medicine, Max Rady College
of Medicine, Rady Faculty of
Health Sciences, University
of Manitoba, Health Sciences
Center, GF-533, 820
Sherbrook Street, Winnipeg,
MB R3A 1R9, Canada.
rmarrie@hsc.mb.ca**Ruth Ann Marrie,**
Department of Internal
Medicine, Max Rady College
of Medicine, Rady Faculty of
Health Sciences, University
of Manitoba, Winnipeg,
CanadaDepartment of Community
Health Sciences, Max Rady
College of Medicine, Rady
Faculty of Health Sciences,
University of Manitoba,
Winnipeg, Canada**Qier Tan,**
Okechukwu Ekuma,
Manitoba Centre for Health
Policy, Max Rady College of
Medicine, Rady Faculty of
Health Sciences, University
of Manitoba, Winnipeg,
Canada**James J Marriott,**
Department of Internal
Medicine, Max Rady College
of Medicine, Rady Faculty of

participants from the sole MS Clinic in Manitoba during the period April 1, 2016 to March 31, 2017. This constituted our linked cohort for further study.

Ethics approval was provided by the Health Research Ethics Board of the University of Manitoba. Approvals for data access were granted by the Health Information Privacy Committee and the Winnipeg Regional Health Authority.

We created candidate indicators for ‘ever smoked’. These included ≥ 1 dispensation of a therapy used for smoking cessation (varenicline, nicotine products, bupropion) captured using DINs; any diagnosis code for tobacco use;⁴ and any diagnosis code for chronic obstructive pulmonary disease (COPD) (Table e1). For COPD we tested a narrow definition limited to individuals aged ≥ 35 years (indicator 1), and two broader indicator without age limits which added the less specific ICD-9-CM/ICD-10-CA 490, 493/J40, J45 diagnosis codes (Table e1).⁵ For each indicator we tested one-year, two-year, five-year and lifetime lookback periods. We compared smoking status based on these indicators, or combinations of those indicators (hereafter ‘definitions’), to smoking status from the MS Clinic questionnaires (gold standard) using two by two contingency tables to estimate sensitivity, specificity, PPV and negative predictive value (NPV). Statistical analyses used SAS V9.4 (SAS Institute Inc., Cary, NC).

Results

After linkage of administrative and smoking status data for the period April 1, 2016 to March 31, 2017, we included 1,320 individuals with MS, of mean age at diagnosis 36.5 (10.4) years. Overall, 788 (59.7%) of individuals had ever smoked (Table e2).

The sensitivity of most candidate indicators and definitions was low ($<32\%$) although it reached 74.5% for the lifetime definitions with broad COPD indicators (Table 1); nicotine products were not detected likely because these are available over the counter. All of the indicators except the two definitions that included the broad COPD indicators had a high specificity, ranging from 86.8–99.8%, and PPV ranging from 58.33% to 98.86%.

Discussion

We tested individual indicators and their combinations to determine ‘ever smoking’ status using administrative data. The PPVs for most combinations were high, exceeding the recommended threshold of 70%,⁶ but sensitivities were low; suggesting that while we

can be confident that individuals classified by these indicators have ever smoked, we will have a high frequency of false negatives. The more sensitive, broad definitions had lower specificities and PPVs $<70\%$. An Australian study compared smoking status among hospitalized patients identified based on ICD-10-Australian modification codes as compared to self-report, and found that sensitivities ranged from 45–74%, and specificities ranged from 94–98% depending on the algorithms used. Sensitivity improved and specificity decreased as they moved from including the most recent hospitalization to a five-year lookback period.⁴ A study in the United States using Medicare data in an rheumatoid arthritis population found that an algorithm employing only diagnosis and procedure codes (e.g. smoking cessation counseling) over a one-year lookback period had a sensitivity of only 9.8% when compared to self-reported smoking status.¹ Incorporating prescription claims and any available data (like a lifetime lookback period) improved sensitivity to 27.9%, with a specificity and PPV of 100%. Notably, smokers identified using the claims-based approach had a longer smoking history than those who were not. A second American study which compared performance of ICD-9 tobacco use codes recorded in an electronic medical record to smoking status based on medical records review found a sensitivity of 32% and specificity of 100%.²

This study has limitations. Several of the tobacco diagnosis codes used required 4 or 5 digits in ICD-9-CM; they were too non-specific for use at the 3-digit level as used in Manitoba’s physician claims. It is possible that if we had access to 4- or 5-digit ICD codes in physician claims we could have improved the sensitivity of our case definition while retaining specificity. We also lacked another population for external validation of our findings.

We showed that ever smokers can be identified with a high degree of specificity, supporting studies focused specifically on examining outcomes in smokers with MS, although some individuals would be missed. Our smoking status definitions would only partially control for confounding due to smoking because of the low sensitivity. Additional strategies to capture health behaviors in administrative data are needed.

Acknowledgements

The authors acknowledge the Manitoba Centre for Health Policy for use of the Manitoba Population Research Data Repository under project #2019-032 (HIPC #2018/2019-66). The results and conclusions presented are those of the authors and no official endorsement by the

Table 1. Performance of indicators and administrative case definitions for smoking status as compared to self-report.

	Sensitivity	Specificity	PPV	NPV
<i>One Year</i>				
Varenicycline	3.17 (2.06, 4.65)	99.81 (98.96, 100)	96.15 (80.36, 99.90)	41.04 (38.34, 43.77)
Bupropion	4.70 (3.33, 6.41)	97.74 (96.09, 98.83)	75.51 (61.13, 86.66)	40.91 (38.19, 43.67)
Tobacco diagnosis	0.25 (0.03, 0.91)	99.62 (98.65, 99.95)	50.00 (6.76, 93.24)	40.27 (37.61, 42.98)
COPD (indicator 1) ^a	2.54 (1.56, 3.89)	99.62 (98.65, 99.95)	90.91 (70.84, 98.88)	40.83 (38.14, 43.56)
COPD (indicator 2) ^b	5.20 (3.60, 6.99)	97.93 (96.33, 98.96)	78.85 (65.30, 88.94)	41.09 (38.36, 43.85)
COPD (indicator nition 3) ^c	10.15 (8.13, 12.48)	93.42 (90.97, 95.38)	69.57 (60.29, 77.80)	41.24 (38.45, 44.08)
Varenicycline OR bupropion OR tobacco diagnosis	7.99 (6.20, 10.11)	97.18 (95.39, 98.41)	80.77 (70.27, 88.82)	41.63 (38.87, 44.43)
Varenicycline OR bupropion OR tobacco diagnosis OR COPD (definition 1)	10.03 (8.02, 12.34)	96.99 (95.16, 98.27)	83.16 (74.10, 90.06)	42.12 (39.34, 44.94)
Varenicycline OR bupropion OR tobacco diagnosis OR COPD (definition 2)	12.56 (10.33, 15.08)	95.30 (93.14, 96.94)	79.84 (71.69, 86.51)	42.39 (39.57, 45.25)
Varenicycline OR bupropion OR tobacco diagnosis OR COPD (definition 3)	16.88 (14.33, 19.68)	90.98 (88.22, 93.27)	73.48 (66.42, 79.75)	42.49 (39.60, 45.42)
<i>Two Year</i>				
Varenicycline	5.20 (3.76, 6.99)	99.81 (98.96, 100)	97.62 (87.43, 99.94)	41.55 (38.83, 44.31)
Bupropion	5.84 (4.31, 7.71)	96.99 (95.16, 98.27)	74.19 (61.50, 84.47)	41.02 (38.28, 43.79)
Tobacco diagnosis	0.51 (0.14, 1.29)	99.62 (98.65, 99.95)	66.67 (22.28, 95.67)	40.33 (37.67, 43.04)
COPD (indicator 1) ^a	3.68 (2.48, 5.24)	99.25 (98.09, 99.79)	87.88 (71.80, 96.60)	41.03 (38.32, 43.77)
COPD (indicator 2) ^b	7.87 (6.09, 9.97)	95.86 (93.81, 97.39)	73.81 (63.07, 82.80)	41.26 (38.50, 44.07)
COPD (indicator 3) ^c	13.83 (11.50, 16.44)	89.85 (86.96, 92.28)	66.87 (59.08, 74.04)	41.31 (38.46, 44.21)
Varenicycline OR bupropion OR tobacco diagnosis	11.17 (9.05, 13.58)	95.86 (93.81, 97.39)	80.00 (71.30, 87.02)	42.15 (39.35, 44.99)
Varenicycline OR bupropion OR tobacco diagnosis OR COPD (definition 1)	14.21 (11.85, 16.85)	95.30 (93.14, 96.94)	81.75 (74.25, 87.83)	42.86 (40.02, 45.73)
Varenicycline OR bupropion OR tobacco diagnosis OR COPD (definition 2)	18.02 (15.40, 20.88)	92.29 (89.69, 94.41)	77.60 (70.86, 83.42)	43.18 (40.28, 46.12)
Varenicycline OR bupropion OR tobacco diagnosis OR COPD (definition 3)	22.97 (20.07, 26.07)	86.28 (83.96, 89.09)	71.26 (65.27, 76.74)	43.06 (40.06, 46.09)
<i>Five Year</i>				
Varenicycline	11.04 (8.94, 13.44)	99.81 (98.96, 100)	98.86 (93.83, 99.97)	43.10 (40.31, 45.92)
Bupropion	8.88 (6.99, 11.09)	96.43 (94.48, 97.84)	78.65 (68.69, 86.63)	41.67 (38.90, 44.49)
Tobacco diagnosis	0.89 (0.36, 1.82)	99.06 (97.82, 99.69)	58.33 (27.67, 84.83)	40.29 (37.62, 43.01)
COPD (indicator 1) ^a	5.46 (3.98, 7.28)	98.87 (97.56, 99.59)	87.76 (75.23, 95.37)	41.38 (38.66, 44.15)
COPD (indicator 2) ^b	13.83 (11.50, 16.44)	90.23 (87.38, 92.61)	67.70 (59.89, 74.85)	41.42 (38.56, 44.31)

(continued)

Table 1. Continued.

	Sensitivity	Specificity	PPV	NPV
COPD (indicator 3) ^c	21.19 (18.39, 24.22)	82.71 (79.22, 85.83)	64.48 (58.32, 70.30)	41.47 (38.49, 44.50)
Varenicycline OR bupropion OR tobacco diagnosis	19.29 (16.59, 22.22)	95.30 (93.14, 96.94)	85.88 (79.86, 90.65)	44.36 (41.45, 47.29)
Varenicycline OR bupropion OR tobacco diagnosis OR COPD (definition 1)	22.97 (20.07, 26.07)	94.36 (92.05, 96.16)	85.78 (80.33, 90.20)	45.27 (42.31, 48.25)
Varenicycline OR bupropion OR tobacco diagnosis OR COPD (definition 2)	29.44 (26.28, 32.76)	86.65 (83.47, 89.43)	76.57 (71.39, 81.22)	45.33 (42.24, 48.45)
Varenicycline OR bupropion OR tobacco diagnosis OR COPD (definition 3)	35.53 (32.19, 38.99)	79.32 (75.63, 82.69)	71.79 (67.05, 76.21)	45.38 (42.14, 48.64)
<i>Lifetime</i>				
Varenicycline	11.04 (8.94, 13.44)	99.81 (98.96, 100)	98.86 (93.83, 99.97)	43.10 (40.31, 45.92)
Bupropion	15.36 (12.91, 18.07)	92.86 (90.33, 94.90)	76.10 (68.70, 82.50)	42.55 (39.68, 45.45)
Tobacco diagnosis	2.92 (1.86, 4.35)	97.93 (96.33, 98.96)	67.65 (49.47, 82.61)	40.51 (37.82, 43.25)
COPD (indicator 1) ^a	10.03 (8.02, 12.34)	95.86 (93.81, 97.39)	78.22 (68.90, 85.82)	41.84 (39.05, 44.66)
COPD (indicator 2) ^b	62.18 (58.69, 65.58)	46.80 (42.50, 51.15)	63.39 (59.88, 66.79)	45.52 (41.29, 49.80)
COPD (indicator 3) ^c	67.64 (64.25, 70.90)	37.59 (31.82, 40.14)	61.62 (58.28, 64.87)	43.96 (39.34, 48.65)
Varenicycline OR bupropion OR tobacco diagnosis	25.76 (22.74, 28.97)	90.60 (87.80, 92.94)	80.24 (74.79, 84.96)	45.17 (42.16, 48.22)
Varenicycline OR bupropion OR tobacco diagnosis OR COPD (definition 1)	31.98 (28.73, 35.36)	86.84 (83.67, 89.60)	78.26 (73.35, 82.64)	46.29 (43.16, 49.44)
Varenicycline OR bupropion OR tobacco diagnosis OR COPD (definition 2)	70.43 (67.11, 73.60)	44.17 (39.90, 48.51)	65.14 (61.83, 68.34)	50.21 (45.59, 54.84)
Varenicycline OR bupropion OR tobacco diagnosis OR COPD (definition 3)	74.49 (71.30, 77.50)	35.90 (31.82, 40.14)	63.25 (60.06, 66.36)	48.72 (43.67, 53.79)

^a≥1 diagnosis code (J41-J44, 490–492, 496) and aged ≥35 years.
^b≥1 diagnosis code (J40-J44, 490–492, 496).
^c≥1 diagnosis code (J40-J45, 490–493, 496).
COPD: chronic obstructive pulmonary disease; PPV: positive predictive value; NPV: negative predictive value.

Manitoba Centre for Health Policy, Manitoba Health, Winnipeg Regional Health Authority, or other data providers is intended or should be inferred.


Declaration of conflicting interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Ruth Ann Marrie receives research funding from CIHR, the MS Society of Canada, Research Manitoba, the CMSC, National MS Society, US Department of Defense, and Crohn's and Colitis Canada. She is supported by the Waugh Family Chair in Multiple Sclerosis. She is a co-investigator on studies funded by Biogen Idec and Roche. Qier Tan has no conflicts of interest to declare. Okechukwu Ekuma has no conflicts of interest to declare. James Marriott receives research funding from the Multiple Sclerosis Scientific Foundation, Research Manitoba and Roche Canada.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by Research Manitoba.

ORCID iD

Ruth Ann Marrie  <https://orcid.org/0000-0002-1855-5595>

Supplemental material

Supplemental material for this article is available online.

References

1. Desai RJ, Solomon DH and Shadick N, et al. Identification of smoking using Medicare data—a validation study of claims-based algorithms. *Pharmacoepidemiol Drug Saf* 2016; 25: 472–475.
2. Wiley LK, Shah A, Xu H, et al. ICD-9 tobacco use codes are effective identifiers of smoking status. *J Am Med Assoc* 2013; 20: 652–658.
3. Marrie RA, Tan Q and Ekuma O, et al. Development and internal validation of a disability algorithm for multiple sclerosis in administrative data. *Front Neurol* 2021; 12: 754144.
4. Havard A, Jorm LR and Lujic S. Risk adjustment for smoking identified through tobacco use diagnoses in hospital data: a validation study. *PLoS One* 2014; 9: e95029.
5. Gershon AS, Wang C and Guan J, et al. Identifying individuals with physician diagnosed COPD in health administrative databases. *COPD* 2009; 6: 388–394.
6. Carnahan RM. Mini-Sentinel's systematic reviews of validated methods for identifying health outcomes using administrative data: summary of findings and suggestions for future research. *Pharmacoepidemiol Drug Saf* 2012; 21: 90–99.