

Administrative data for observational research in multiple sclerosis: Opportunities and challenges

Ruth Ann Marrie  and Kyla McKay

People with multiple sclerosis (MS) have high rates of health care use, including hospitalizations, one of the most costly forms of care. Relatively little is known about the frequency of readmissions after hospitalizations in persons with MS, or factors associated with readmission. This is important because readmissions are costly, and some are avoidable. Although readmissions are influenced by the patient's characteristics, social networks, and the care delivered, they are viewed by some as measures of quality of care.¹

In this issue of *Multiple Sclerosis Journal (MSJ)*, Schorr et al. used the 2013 National Readmissions Database (United States) to examine readmissions for depression or suicide attempt among individuals admitted with MS as compared to asthma or rheumatoid arthritis (RA). They identified these chronic conditions and comorbid psychiatric and substance use disorders using a single hospital claim. The readmission rate for depression was higher after an admission for MS than for asthma (hazard ratio (HR) 1.37; 1.00–1.86) or RA (HR 4.68; 1.60–13.62). Clinically, this indicates a need to improve the identification and effective treatment of psychiatric conditions in people with MS.

This study also highlights the opportunities and challenges related to administrative data. Administrative (health claims) data are generated through delivery of and reimbursement for health care services.² In the United States, a mixture of administrative data sources exists, including public (e.g. Medicare, Veteran's Health Administration) and private (e.g. commercial). Typically, these data sources include a unique identifier, demographic information such as sex, date of birth, region of residence, date of service and diagnostic and procedure codes. As compared to primary data collection, administrative data are accessible, available at relatively low cost and bear minimal patient burden.² Furthermore, in regions with universal health systems, they are population-based, limiting selection

bias, and offering longitudinal follow-up. Sample sizes are large, supporting the identification and study of rare events and conditions.

Despite the advantages of administrative data, limitations also need to be considered carefully. First, sociodemographic information such as race and ethnicity may not be collected, thereby limiting the ability to understand or account for social determinants of health. Second, detailed clinical information such as disability status is lacking, although it is sometimes possible to mitigate this issue through linkage to clinical data sources. Third, since the administrative data are collected for reimbursement, they reflect the care for which insurers were billed.³ Therefore when multiple conditions were assessed or multiple services provided, the conditions which are coded may be those for which reimbursement is highest. Comorbid conditions may not be completely captured in hospital claims due to coding biases.⁴ Coding practices may change over time, creating temporal trends which do not reflect true changes in epidemiology or clinical practice.⁵

Finally, the validity of administrative data needs to be considered, as they are not collected for research purposes. Therefore, studies that develop and validate case definitions are critical, particularly for conditions that are more complex to diagnose such as MS and RA. For example, a single diagnosis code may indicate that the diagnosis was being "ruled out" rather than being confirmed. Studies in the United States, Canada, and Sweden suggest that the occurrence of at least three diagnosis codes in hospital or physician claims in any combination, is needed to achieve high positive and negative predictive values for MS.⁶ This study identified all conditions, including MS, based on one hospital claim. While specificity is typically high based on hospital claims, this is not universal, and positive predictive values may be quite low for some conditions.⁷ Thus, it is likely that some individuals without MS

Multiple Sclerosis Journal

2022, Vol. 28(1) 3–6

DOI: 10.1177/

13524585211055787

© The Author(s), 2021.



Article reuse guidelines:
sagepub.com/journals-
permissions

Correspondence to:

RA Marrie
Health Sciences Center,
University of Manitoba, GF-
533, 820 Sherbrook Street,
Winnipeg, MB R3A 1R9,
Canada.
rmarrie@hsc.mb.ca

Ruth Ann Marrie
Departments of Internal
Medicine and Community
Health Sciences, Max Rady
College of Medicine, Rady
Faculty of Health Sciences,
University of Manitoba,
Winnipeg, MB, Canada

Kyla McKay
Department of Clinical
Neuroscience, Karolinska
Institutet, Stockholm,
Sweden

were captured in the MS cohort. Similarly, identifying RA based on one hospital claim is challenging. For example, only 59% of individuals identified based on a single RA claim in the Danish National Hospital Patient Register were confirmed to have RA after medical records review.⁸ The accurate identification of psychiatric disorders is also challenging, and case definitions typically suffer from low sensitivity.⁹ Thus, prevalence estimates based on administrative data generally underestimate the true burden of psychiatric conditions. This low sensitivity (or high proportion of false negatives) is due to a multitude of factors including the persistent stigma attached to mental health that leads persons to avoid care;¹⁰ mental health care which is not always captured in administrative data (for instance, primary care or private counselling), and coding biases in which conditions such as depression are under-coded in the presence of another condition.⁴

As illustrated in Figure 1, the implications of imperfect case definitions vary depending on the performance characteristics of the definition, the populations in whom they are applied, and the study purpose. In Panel a, a case definition with high sensitivity and specificity performs reasonably well but misclassification is

greater when the condition of interest is very rare or very common. In Panel b, modifying the case definition to reduce the specificity markedly worsens performance, leading to a high proportion of false positives. This misclassification is likely to reduce the ability to detect differences between groups with and without the condition. In Panel c, modifying the case definition to reduce the sensitivity while retaining high specificity also worsens performance when estimating prevalence due to false negatives (missed cases). However, the positive predictive value remains high, so one can be confident they have identified people with the condition of interest.

Studies developing high performing case definitions for chronic disease are critical to realizing the full potential of administrative data, and this work needs to be repeated when changes in diagnostic coding systems occur, or when case definitions are applied in jurisdictions with differing health systems or coding practices. While administrative data remain a valuable resource for large-scale epidemiological investigations, they contain potential biases and other limitations which should be clearly acknowledged in studies which employ them.

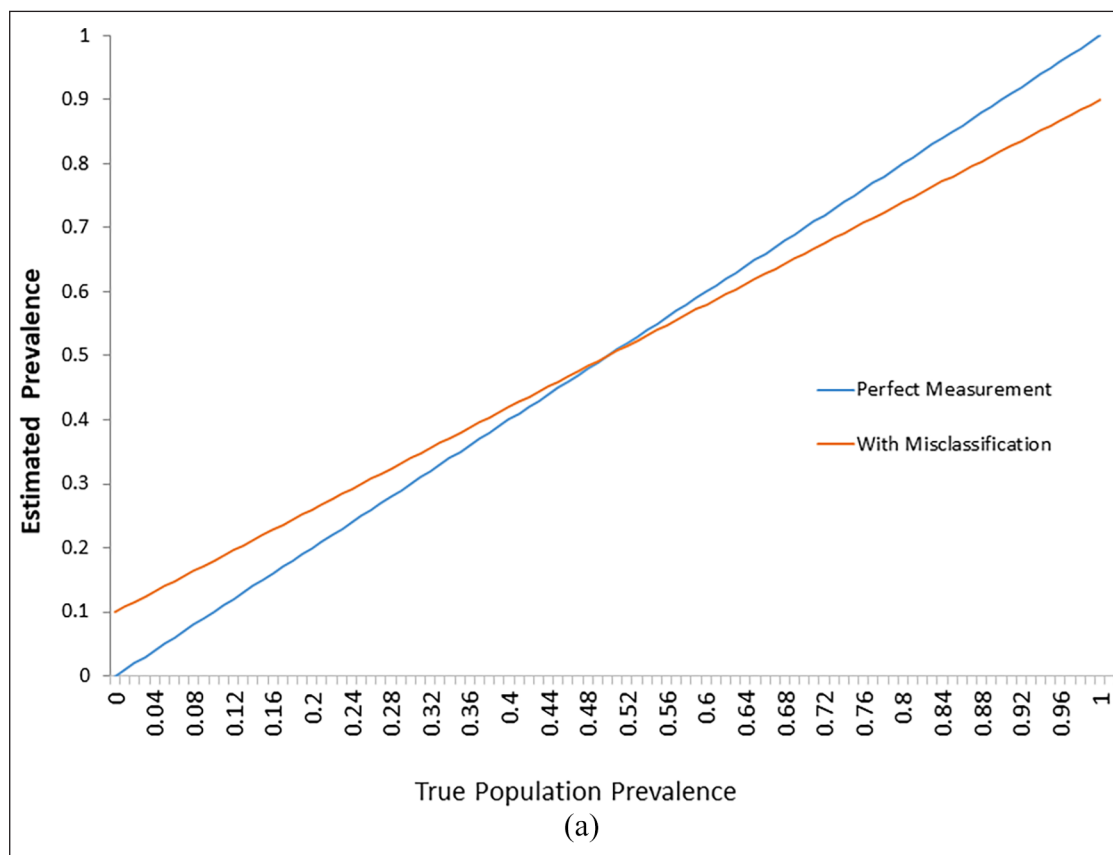


Figure 1. (Continued)

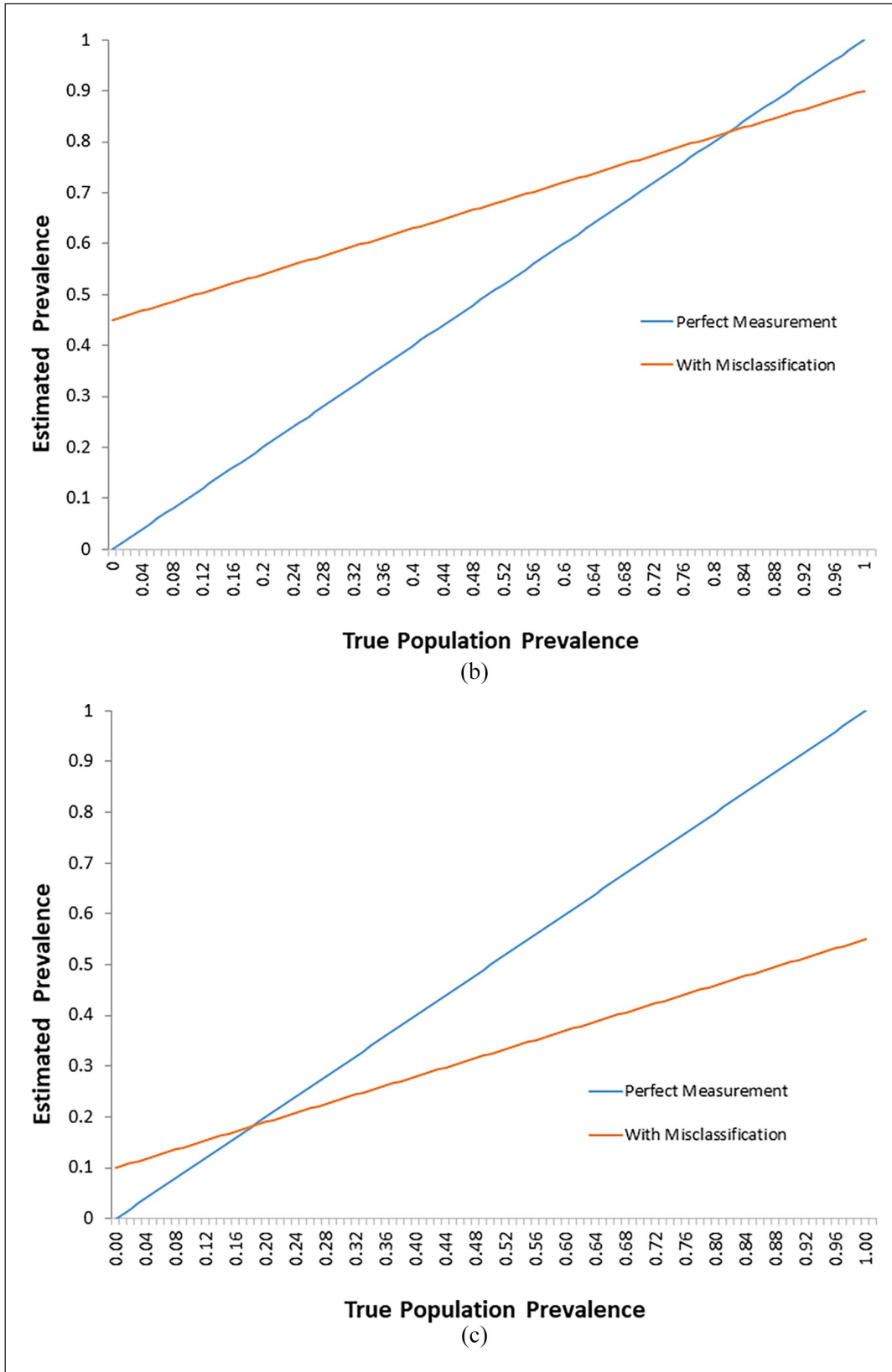


Figure 1. Misclassification bias: Comparison between estimated and true population prevalence: (a) Case definition sensitivity: 0.90, specificity 0.90. (b) Case definition sensitivity: 0.90, specificity: 0.55. (c) Case definition sensitivity: 0.55, specificity 0.99.

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, U.S. Department of Defense, and Crohn's and Colitis Canada. She is a co-investigator on studies funded by Biogen Idec and Roche.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: KAM receives research funding from the Swedish Research Council for Health, Working Life and Welfare (Forte). Ruth Ann Marrie is supported by the Waugh Family Chair in Multiple Sclerosis. No funding was received for this editorial.

ORCID iD

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

References

1. Fischer C, Lingsma HF, Marang-van de Mheen PJ, et al. Is the readmission rate a valid quality indicator? A review of the evidence. *PLoS ONE* 2014; 9: e112282.
2. Iezzoni LI. Assessing quality using administrative data. *Ann Intern Med* 1997; 127: 666–674.
3. Virnig BA and McBean M. Administrative data for public health surveillance and planning. *Ann Rev Publ Health* 2001; 22: 213–230.
4. Peng M, Southern DA, Williamson T, et al. Under-coding of secondary conditions in coded hospital health data: Impact of co-existing conditions, death status and number of codes in a record. *Health Inform J* 2017; 23(4): 260–267.
5. Li L, Binney LE, Luengo-Fernandez R, et al. Temporal trends in the accuracy of hospital diagnostic coding for identifying acute stroke: A population-based study. *Eur Stroke J* 2020; 5: 26–35.
6. Culpepper WJ, Marrie RA, Langer-Gould A, et al. Validation of an algorithm for identifying MS cases in administrative health claims datasets. *Neurology* 2019; 92: e1016–e28.
7. Quan H, Li B, Saunders LD, et al. Assessing validity of ICD-9-CM and ICD-10 administrative data in recording clinical conditions in a unique dually coded database. *Health Serv Res* 2008; 43(4): 1424–1441.
8. Pedersen M, Klarlund M, Jacobsen S, et al. Validity of rheumatoid arthritis diagnoses in the Danish National Patient Registry. *Eur J Epidemiol* 2004; 19: 1097–1103.
9. Marrie RA, Fisk JD, Yu BN, et al. Mental comorbidity and multiple sclerosis: validating administrative data to support population-based surveillance. *BMC Neurol* 2013; 13: 16.
10. Schomerus G and Angermeyer MC. Stigma and its impact on help-seeking for mental disorders: what do we know. *Epidemiol Psychiatr Soc* 2008; 17(1): 31–37.