Family Practice, 2019, 804–807
doi:10.1093/fampra/cmy130

CASFM Methods Briefs

Recognizing misclassification bias in research and medical practice

Anh Pham[a], Michael Cummings[b], Cliff Lindeman[b], Neil Drummond[b] and Tyler Williamson[c,*]

[a]Department of Family Medicine, University of Calgary, Calgary, Alberta, Canada, [b]Department of Family Medicine, University of Alberta, Edmonton, Alberta, Canada and [c]Department of Community Health Sciences, University of Calgary, Calgary, Alberta, Canada.

*Correspondence to Tyler Williamson, Department of Community Health Sciences, University of Calgary, TRW Building 3rd Floor, 3280 Hospital Drive NW, Calgary, Alberta T2N 4Z6, Canada; E-mail: tyler.williamson@ucalgary.ca

Key words: Accuracy, case definition, epidemiology, errors, information bias, misclassification.

Introduction

Systematic errors are common in health care research. Although methods for recognizing biases and adjusting for them are well discussed in many textbooks (1,2), a lack of familiarity with these concepts can easily lead researchers and clinicians to draw incorrect conclusions when interpreting research. In order to increase awareness of the challenges systematic errors cause, this paper will discuss one class of errors, misclassification bias, and will provide examples from commonly-used health care data.

Misclassification bias

Misclassification bias is a systematic error that can occur at any stage in the research process. It occurs when an individual is assigned to a different category than the one to which they should be assigned (1). For example, if a patient appears to be non-hypertensive because of medication-controlled blood pressure, resulting in systolic and diastolic measures that are within the 'normal range', this may constitute an incorrect classification.

Two types of misclassification bias are of particular importance when studying disease status.

1. *Misclassification of exposure* arises when errors or biases occur during collection of exposure data (3). For example, people who have dementia may be less likely to remember specific risk factors that they experienced earlier in life. They may subsequently self-report as being unexposed to these risk factors when, in fact, they have been exposed. At the other end of the spectrum, populations that have experienced a specific, memorable exposure may be more likely to recall this event. For example, hypertensive patients who have suffered a heart attack may be more likely to recall their medical history of hypertension. They may then be more likely to insist on a relationship between hypertension and other diseases later in their lives than people who have not experienced a cardiac event. In both instances, these *recall biases* can lead to a misclassification of exposure (4).
2. *Misclassification of outcome* derives from errors or biases in the collection of outcome data, such as disease status (1). As an example, patients who do not have a family history of dementia may be tested less often than those who do. This may result in a greater proportion of missed dementia cases in populations without a family history of dementia than those with it.

Accuracy of case definitions

In the context of this paper, a case definition is defined as a set of rules that identify if an individual has a disease or condition. A case definition's utility is determined by its sensitivity and specificity in comparison to a reference standard, which are often derived from chart reviews by qualified persons (5–7), numerical test criteria or clinician interviews. For a case definition to be epidemiologically meaningful, it must reach a certain level of accuracy quantified by its 'sensitivity' (the probability that a case definition correctly classifies individuals who have the disease) and 'specificity' (the probability that a case definition correctly classifies individuals who do not have the disease) (4). For epidemiological research, an accepted degree of sensitivity and specificity is 70% or greater (7). A high specificity and sensitivity demonstrate that the case definition correctly identifies high proportions of true cases and non-cases in the cohort, and so misclassification of outcome will be minimized.

As a hypothetical example, suppose a study includes 100 cases and 100 non-cases of a given disease, and assume that the case definition in question has a 90% sensitivity and 80% specificity. As

Table 1. Example calculation of sensitivity and specificity for a cohort with 100 patients who have a disorder and 100 who do not

	Has disorder (n = 100)	Does not have disorder (n = 100)
Meets case definition criteria	True Positive (TP) = 90	False Positive (FP) = 20
Does not meet case definition criteria	False Negative (FN) = 10	True Negative (TN) = 80
Accuracy of the case definition	Sensitivity = TP/(TP+FN) = 90%	Specificity = TN/(TN+FP) = 80%

illustrated in Table 1, in 100 individuals with this disease, 90 meet the case-definition criteria of having the disease (the 'true positives'); in 100 persons who do not have the disease, 80 do not meet the case-definition criteria (the 'true negatives'). The remaining patients (the 'false positives' and 'false negatives') are incorrectly identified (8). High specificity and sensitivity demonstrate that the case definition correctly identifies high proportions of true cases and non-cases in the cohort, and so misclassification of outcome will be minimized.

Clinically, two other metrics are generally more relevant in evaluating case-definition validity, namely positive and negative predictive value. These are influenced by disease prevalence: the higher the prevalence of a condition in a group of people, the higher the probability that any individual member of the group will have the condition. Given that the purpose of this paper is to study misclassification specifically, we shall look only at sensitivity and specificity, as they describe the ability of a case definition to correctly identify cases and non-cases, regardless of the prevalence of a disease within a population (9). Further reading on how disease case definitions have been implemented and validated may be found elsewhere (7,10–13).

Examples of misclassification bias

Electronic medical record (EMR) databases are rich sources of information, enabling researchers to conduct observational studies that assess the occurrence and management of chronic diseases over time. Furthermore, EMR data allows assessment of routine care and clinical behaviour in a natural setting. In an epidemiological context, processed EMR data can be used to (i) calculate disease incidence rates and prevalence; (ii) identify at-risk groups and (iii) estimate the burden of disease in a population (14). For the purposes of this paper, these data can also be used to demonstrate how misclassification bias can lead to erroneous research conclusions.

In Canada, the Canadian Primary Care Sentinel Surveillance Network (CPCSSN) is the only pan-Canadian platform for chronic disease surveillance in primary care settings. As of 2018, CPCSSN extracts EMR data from eight provinces or territories, and creates a longitudinal data set that is standardized and stored in a secure, de-identified data repository (15).

To illustrate the effects of misclassification bias, validated CPCSSN case definitions for dementia and hypertension have been applied to CPCSSN EMR data for patients who are older than 60 (7). The result is a 'two-by-two table' of disease occurrence (Table 2). The case definition for dementia used here is excellent, with sensitivity of 96.8% and specificity of 98.1% (7). For hypertension, the values are slightly lower with 85% sensitivity and 93% specificity (7).

The effect of misclassification bias on these data can be illustrated through the relative risk, or risk ratio (RR), a measure for describing the association between the two diseases. The risk ratio is given as

$$RR = \left(\frac{a}{a+b}\right) \Big/ \left(\frac{c}{c+d}\right)$$

where the numerator is the proportion of individuals with the second disease among those who have the first disease, and the denominator is the proportion of individuals who have the second disease among those who do not have the first disease (Table 2). For the data given in Table 2, the risk ratio for dementia is 1.79. That is, a patient with hypertension is 1.79 times more likely to develop dementia than a patient without hypertension. It should be noted that this calculation of risk is an oversimplification and has been used to more clearly illustrate the effects of misclassification bias. In actuality, risk estimation should be considered in a more complete context, as individuals may have been exposed to multiple diseases and additional confounding factors may be present.

Table 2. A two-by-two table of the number of patients with or without hypertension and dementia as identified by CPCSSN case definitions

	Dementia		
Hypertension	Yes	No	Total
Yes	1999[a]	22,889[b]	24,888
No	1542[c]	32,740[d]	34,282
Total	3541	55,629	59,170

[a, b, c, d]Represent the variables in the risk ratio equation.

Beyond the fact that this is a simplified example, the case definitions for dementia and hypertension are imperfect—their sensitivities and specificities are less than 100%. Therefore, some patients will be misclassified as having or not having these diseases and the estimated risk ratio will be biased. The amount of bias can be demonstrated by examining the two-by-two table for a hypothetical scenario where the case definition for dementia is perfect (i.e. no misclassification) while the case definition for hypertension remains unchanged. In this case, the hypothetical number of patients with or without each disease is given in Table 3. While, in practice, perfectly accurate case definitions do not exist, nonetheless, we will use one here to illustrate that care must still be taken when interpreting results generated with highly accurate, but imperfect, case definitions.

The data in Tables 2 and 3 demonstrate that 994 (i.e. 3541–2547), or ~2%, of patients in Table 2 appear to be misclassified as having dementia when, in fact, they do not. Furthermore, the hypothetical risk ratio for dementia, comparing those with hypertension to those without, is RR = 2.36. This implies that the previous estimate of RR = 1.79, calculated using the imperfect dementia case definition, substantially underestimates the true association between hypertension and dementia. This example illustrates that we need to carefully consider the effects of misclassification bias on study findings, as these biases may significantly impact the interpretation of results.

The effects of misclassification bias become even more apparent when a case definition with sensitivity and specificity that are on the edge of acceptability for epidemiological research is used. Table 4 contains another two-by-two table based on a dementia case definition with sensitivity and specificity of 70%. Not only is the number of patients with dementia greatly overestimated, the association between hypertension and dementia is nearly absent, as the risk ratio is close to 1 (RR = 1.05). In other words, the risk of developing dementia among hypertensive patients is underestimated because

a large number of patients, both with and without dementia, have been misclassified. Based on the low accuracy of the case definition, one could conclude that being hypertensive has a negligible effect on developing dementia, whereas more accurate case definitions like those used to calculate the values given in Tables 2 and 3 indicate hypertension and dementia are associated.

In the preceding examples, the context was simplified to provide a clear explanation about sources of misclassification bias. More complex examples may be found elsewhere (3,16).

Preventing misclassification bias

The effects of misclassification bias are not restricted to case-definition research; they may also influence interpretation of laboratory results or other diagnostic procedures. Regardless of the application, if sensitivity and specificity are less than 100%, some degree of misclassification will occur and may have a profound impact on clinical or research conclusions. Perfect tests and case definitions do not exist due to various types of error, bias and effects that occur simply by random chance (8). Moreover, the more sensitive a case definition or a diagnostic method is, the less specific it will tend to be (9). It is therefore vital for researchers and practitioners to exercise careful judgement as to whether reducing false negatives or false positives are more important given the purpose of a calculation. For instance, screening for disease prevalence in a primary care setting may need a highly sensitive test to avoid missing cases. Conversely, a randomized controlled trial may require highly specific criteria to ensure that only true disease cases are included in the study, accepting that many people who do have the disease may be rejected from the cohort.

Correcting systematic misclassification errors that occurred during data collection may not be possible when analysing secondary data sources. Therefore, care should be taken to minimize the likelihood of misclassification during data collection. This can be accomplished by having a detailed, straightforward and consistent case definition, strictly following diagnosis guidelines, and minimizing measurement errors by selecting more accurate equipment, tests or medical examination procedures. On a cohort level, one may also want to repeat a specific test or case definition to obtain the most accurate result (17). Fortunately, many errors encountered when classifying data post-collection can be corrected by applying alternative classifying tools or analysis strategies; however, this may require external resources and expertise to accomplish (18,19).

Table 3. Hypothetical patient numbers for a perfect case definition for dementia

	Dementia		
Hypertension	Yes	No	Total
Yes	1608	23,280	24,888
No	939	33,343	34,282
Total	2547	56,623	59,170

Table 4. Hypothetical patient numbers for a low sensitivity and specificity case definition for dementia

	Dementia		
Hypertension	Yes	No	Total
Yes	8110	16,778	24,888
No	10,660	23,622	34,282
Total	18,770	40,400	59,170

Conclusion

Using an inaccurate case definition may lead to misclassification bias, and unsuccessfully recognizing and adjusting for misclassification bias may lead to incorrect conclusions about the relationship between an exposure and its outcome. Errors caused by misclassification are common and often problematic. Researchers should not assume that a cohort of identified disease cases has perfect accuracy; instead, they should examine and carefully appraise the methods by which cases and exposure status are determined in order to identify potential misclassification bias and prepare a mitigation strategy. When considering research findings, a critical eye should be focussed on the potential inaccuracy of all classification strategies, including case definitions and test procedures, especially if they are not clearly explained, to avoid inaccurate interpretation.

Declaration

Funding: none.
Ethical approval: none.
Conflict of interest: none.

References

1. Szklo M, Nieto FJ. *Understanding Lack of Validity: Bias, in Epidemiology: Beyond the basis*. Burlington, MA: Jones&Barlett Learning, 2014.
2. Rothman, KJ, Sander G, Lash TL. Validity in Epidemiologic Studies. In: Rothman KJ, Sander G, Lash TL (eds). *Modern Epidemiology*. Philadelphia, PA: Lippincott Williams & Wilkins, 2012, pp. 128–147.
3. Szklo M, Nieto FJ. *Information Bias, in Epidemiology Beyond the Basics*. Burlington, MA: Jones & Barlett Learning, 2014, pp. 116–132.
4. Aschengrau A, Seage G. *Screening in Public Health Practice, in Essentials of Epidemiology in Public Health*. Burlington, MA: Jones & Bartlett Learning, 2014.
5. Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and using sensitivity, specificity and predictive values. *Indian J Ophthalmol* 2008; 56: 45–50.
6. Kadhim-Saleh A, Green M, Williamson T, Hunter D, Birtwhistle R. Validation of the diagnostic algorithms for 5 chronic conditions in the Canadian Primary Care Sentinel Surveillance Network (CPCSSN): a Kingston Practice-based Research Network (PBRN) report. *J Am Board Fam Med* 2013; 26: 159–67.
7. Williamson T, Green ME, Birtwhistle R *et al*. Validating the 8 CPCSSN case definitions for chronic disease surveillance in a primary care database of electronic health records. *Ann Fam Med* 2014; 12: 367–72.
8. Weiss NS. Clinical Epidemiology. In: Rothman KJ, Greenland S, Lash TL, (Eds). *Modern Epidemiology*. Philadelphia, PA: Lippincott Williams & Wilkins, 2008.
9. Lalkhen AG, A McCluskey, Clinical tests: sensitivity and specificity. *Continuing Education in Anaesthesia Critical Care & Pain* 2008; 8: 221–223.
10. Krause G, Brodhun B, Altmann D, Claus H, Benzler J. Reliability of case definitions for public health surveillance assessed by Round-Robin test methodology. *BMC Public Health* 2006; 6: 129.
11. Lix L, Singer A, Katz A, Yogendran M, Al-Azazi S. Chronic disease case definitions for electronic medical records: a Canadian validation study. *Int J Popul Data Sci* 2017; 1: 212.
12. Peng M, Chen G, Kaplan G, Lix L, Drummond N, Lucyk K, Garies S, Lowerison M, Weibe S, Quan H. Methods of defining hypertension in electronic medical records: validation against national survey data. *Int J Popul Data Sci* 2017; 1: 039.

13. Coggon D, Martyn C, Palmer KT, Evanoff B. Assessing case definitions in the absence of a diagnostic gold standard. *Int J Epidemiol* 2005; 34: 949–52.
14. Greiver M, Williamson T, Bennett TL *et al.*; Canadian Primary Care Sentinel Surveillance Network. Developing a method to estimate practice denominators for a national Canadian electronic medical record database. *Fam Pract* 2013; 30: 347–54.
15. Birtwhistle RV. Canadian primary care sentinel surveillance network: a developing resource for family medicine and public health. *Can Fam Physician* 2011; 57: 1219–20.
16. Baena A, Garcés-Palacio IC, Grisales H. The effect of misclassification error on risk estimation in case-control studies. *Rev Bras Epidemiol* 2015; 18: 341–56.
17. Aschengrau A, Seage G. *Bias, in Essentials of Epidemiology in Public Health*. Burlington, MA: Jones & Bartlett Learning. 2014, p. 284–287.
18. van Walraven C. Improved correction of misclassification bias with bootstrap imputation. *Med Care* 2018; 56: e39–45.
19. Bakoyannis G, Yiannoutsos CT. Impact of and correction for outcome misclassification in cumulative incidence estimation. *PLoS One* 2015; 10: e0137454.