ERIC Notebook

Second Edition

Sources of Systematic Error or Bias: Information Bias

Second Edition Authors:

Lorraine K. Alexander, DrPH

Brettania Lopes, MPH

Kristen Ricchetti-Masterson, MSPH

Karin B. Yeatts, PhD, MS

Information bias is one type of systematic error that can occur in epidemiologic studies. *Bias* is any systematic error in an epidemiologic study that results in an incorrect estimate of the association between exposure and the health outcome. Bias occurs when an estimated association (risk ratio, rate ratio, odds ratio, difference in means, etc.) deviates from the true measure of association.

Bias is caused by systematic variation, while chance is caused by random variation. The consequence of bias is systematic error in the risk ratio, rate ratio, or odds ratio estimate. Bias may be introduced at the design or analysis phase of a study. We should try to eliminate or minimize bias through study design and conduct.

Major types of systematic error include the following:

- Selection bias
- Confounding bias
- Information bias

In this issue we present information bias. Selection bias and confounding are covered in separate ERIC Notebooks.

Information bias is a distortion in the measure of association caused by a lack of accurate measurements of key study variables. Information bias, also called measurement bias, arises when key study variables (exposure, health outcome, or confounders) are inaccurately measured or classified. Bias in the risk ratio, rate ratio, or odds ratio can be produced even if measured errors are equal between exposed and unexposed or between study participants that have or do not have the health outcome.

Non-differential misclassification

Non-differential misclassification occurs if there is equal misclassification of exposure between subjects that have or do not have the health outcome or if there is equal misclassification of the health outcome between exposed and unexposed subjects. If exposure or the health outcome is dichotomous, then non-differential misclassification causes a bias of the risk ratio, rate ratio, or odds ratio towards the null.

Non-differential misclassification of exposure status

Non-differential misclassification of exposure status in a case-control study occurs when exposure status is

equally misclassified among cases and controls. Non-differential misclassification in a cohort study occurs when exposure status is equally misclassified among persons who develop and persons who do not develop the health outcome.

Non-differential misclassification of health outcome status occurs in a case-control study when the health outcome status is equally misclassified among exposed and unexposed subjects. Non-differential misclassification of the health outcome status occurs in a cohort study when a study subject who develops the health outcome is equally misclassified among exposed and unexposed cohorts.

Effect of non-differential misclassification of exposure

Non-differential misclassification biases the risk ratio, rate ratio, or odds ratio towards the null if the exposure classification is dichotomous, i.e., either exposed or unexposed. If exposure is classified into 3 or more categories, intermediate exposure groups may be biased away from the null, but the overall exposure-response trend will usually be biased towards the null.

Effect of non-differential misclassification of the health outcome

In most cases, non-differential misclassification of the health outcome will produce bias toward the null, i.e. the risk ratio, rate ratio or odds ratio will be biased towards 1.0. If errors in detecting the presence of the health outcome are equal between exposed and unexposed subjects (i.e. sensitivity is less than 100%) but no errors are made in the classification of health outcome status (i.e. specificity is 100%), the risk ratio or rate ratio in a cohort study will not be biased, but the risk difference will be biased towards the null.

Effect of non-differential misclassification of health outcome status

If no errors are made in detecting the presence of the health outcome (i.e. 100% sensitivity), but equal errors are made among exposed and unexposed in the classification of health outcome status (i.e. specificity less than 100%), the risk ratio, rate ratio, and risk difference (as applicable) will be biased towards the null.

Combined errors in both sensitivity and specificity further increase the bias towards the null, but specificity errors produce larger biases overall.

Differential misclassification

Differential misclassification occurs when misclassification of exposure is not equal between subjects that have or do not have the health outcome, or when misclassification of the health outcome is not equal between exposed and unexposed subjects.

Differential misclassification causes a bias in the risk ratio, rate ratio, or odds ratio either towards or away from the null, depending on the proportions of subjects misclassified.

Effect of differential misclassification of exposure or health outcome

Differential misclassification of the exposure or health outcome can bias the risk ratio, rate ratio, or odds ratio either towards or away from the null. The direction of bias is towards the null if fewer cases are considered to be exposed or if fewer exposed are considered to have the health outcome. The direction of bias is away from the null if more cases are considered to be exposed or if more exposed are considered to have the health outcome.

The effect of differential misclassification of the exposure or health outcome can bias the risk ratio, rate ratio, or odds ratio in either direction. The direction of bias is towards null if fewer cases are considered to be exposed or if fewer exposed subjects are considered to have the health outcome. The direction of bias is away from the null if more cases are considered to be exposed or if more exposed subjects are considered to have the health outcome.

Interviewer bias

Interviewer bias is a form of information bias due to:

1. lack of equal probing for exposure history between cases and controls (exposure suspicion bias); or

2. lack of equal measurement of health outcome status between exposed and unexposed (diagnostic suspicion bias)

Solutions:

1. blind data collectors regarding exposure or health outcome status
2. develop well standardized data collection protocols
3. train interviewers to obtain data in a standardized manner
4. seek same information about exposure from two different sources, e.g. index subject and spouse in case-control study

Recall or reporting bias

Recall or reporting bias is another form of information bias due to differences in accuracy of recall between cases and non-cases or of differential reporting of a health outcome between exposed and unexposed.

Cases may have greater incentive, due to their health concerns, to recall past exposures. Exposed persons in a cohort study may be concerned about their exposure and may over-report or more accurately report the occurrence of symptoms or the health outcome.

Solutions:

1. add a case group unlikely to be related to exposure
2. add measures of symptoms or health outcomes unlikely to be related to exposure

Complications in predicting direction of misclassification bias

Misclassification of confounders results in unpredictable direction of bias. Non-differential misclassification of a polychotomous exposure variable (3 or more categories) may result in bias away from null, though this is less likely than bias towards the null.

Non-differential misclassification of a health outcome limited to a loss of sensitivity of detecting the health outcome without any loss in specificity does not bias toward null, whereas a loss of specificity always biases toward the null.

Conclusions

Some inaccuracies of measurement of exposure and health outcome occur in all studies.

If a positive exposure-health outcome association is found and non-differential measurement errors are more likely than differential ones, measurement error itself cannot account for the positive finding since non-differential error nearly always biases towards the null.

Strive to reduce errors in measurement:

1. develop well standardized protocols
2. train interviewers and technicians well
3. perform pilot studies to identify problems with questionnaires and measuring instruments
4. attempt to assess the direction of bias by considering likelihood of non-differential or differential misclassification

Terminology

Information bias: A distortion in the measure of association caused by a lack of accurate measurements of exposure or health outcome status which can result from poor interviewing techniques or differing levels of recall by participants.

Non-differential misclassification: Equal misclassification of exposure between subjects that have or do not have the health outcome, or equal misclassification of the health outcome between exposed and unexposed subjects.

Differential misclassification: Unequal misclassification of exposure between subjects that have or do not have the health outcome, or unequal misclassification of the health outcome between exposed and unexposed subjects.

Practice Questions

Answers are at the end of this notebook

1) Researchers conduct a case-control study. The following table shows the true classification of exposure and the health outcome (Note: these data are hypothetical and typically researchers would not know the true unbiased distribution of exposure and outcome).

	Have health out-come	Do not have health out-come
Exposed	200	210
Unex-posed	340	500

a) Calculate the odds ratio

Now imagine that 50 people with the health outcome were misclassified as being unexposed and 20 people with the health outcome were misclassified as being exposed.

b) Create the corrected 2x2 table

	Have health out-come	Do not have health outcome
Exposed		
Unex-posed		

c) Calculate the odds ratio for the corrected table
d) In which direction was the misclassification bias?

2) Researchers conduct a case-control study of the association between the diet of young children and diagnosis of childhood cancer, by age 5 years. The researchers are worried about the potential for recall bias since parents are being asked to recall what their children generally ate, over a period of 5 years. Which of the following potential control groups would be most likely to reduce the likelihood of recall bias?

a) Parents of children with no known health problems
b) Parents of children with other known, diagnosed serious health problems (aside from childhood cancer)
c) Parents of children with other known, diagnosed minor health problems

References

Dr. Carl M. Shy, Epidemiology 160/600 Introduction to Epidemiology for Public Health course lectures, 1994-2001, The University of North Carolina at Chapel Hill, Department of Epidemiology

Rothman KJ, Greenland S. Modern Epidemiology. Second Edition. Philadelphia: Lippincott Williams and Wilkins, 1998.

Dr. Steve Marshall and Dr. Jim Thomas Epidemiology 710, Fundamentals of Epidemiology course lectures, 2009-2013, The University of North Carolina at Chapel Hill, Department of Epidemiology

Dr. David Richardson Epidemiology 718, Epidemiologic Analysis of Binary Data course lectures, 2009-2013, The

Acknowledgement

The authors of the Second Edition of the ERIC Notebook would like to acknowledge the authors of the ERIC Notebook, First Edition: Michel Ibrahim, MD, PhD, Lorraine Alexander, DrPH, Carl Shy, MD, DrPH, Gayle Shimokura, MSPH and Sherry Farr, GRA, Department of Epidemiology at the University of North Carolina at Chapel Hill. The First Edition of the ERIC Notebook was produced by the Educational Arm of the Epidemiologic Research and Information Center at Durham, NC. The funding for the ERIC Notebook First Edition was provided by the Department of Veterans Affairs (DVA), Veterans Health Administration (VHA), Cooperative Studies Program (CSP) to promote the strategic growth of

Answers to Practice Questions

1.

a) Calculate the odds ratio

Odds ratio = (200*500) / (340*210) =1.4

b) Create the corrected 2x2 table

	Have health out-come	Do not have health outcome
Exposed	230	210
Unex-posed	310	500

c) Calculate the odds ratio for the corrected table

Odds ratio = (230*500) / (310*210) =1.8

d) In which direction was the misclassification bias?

The bias was away from the null (the null value is 1.0).

2. Answer choice b is the best choice. Researchers should aim to have similar recall bias between both the case and control groups. Parents of children who have childhood cancer, which is a serious health problem, are likely to be quite concerned about what may have contributed to the cancer. Thus, if asked by researchers, these parents are likely to think very hard about what their child ate or did not eat in their first years of life. Parents of children with other serious health problems (aside from cancer) are also likely to be quite concerned about any exposure that researchers ask about. Therefore, these parents can be expected to recall exposures in a way that is more comparable with parents of children who have cancer. In contrast, parents of children who have no health problems or parents of children with only minor health problems are less likely to be concerned with carefully recalling any exposures.