

### JOURNAL SERIES:

# Statistics Review Part 7: Case-Control and Cohort Studies

by Claire Seidler, PharmD Candidate 2015, Amanda Margolis, PharmD, MS, BCACP, and Kevin Look, PharmD, PhD

This article describes cohort and case-control studies, differences between the two study types and biases that observational research may be susceptible to.

### Objectives:

1. Define cohort and case-control studies
2. Describe the differences between cohort and case-control studies
3. Describe different types of bias that observational research is susceptible to

Observational studies are often warranted and necessary when a randomized control trial is unethical to perform, when the outcome or condition of interest is rare, or as a hypothesis generating study to determine if future randomized control trials are warranted.<sup>1</sup> Cohort and case-control studies are two commonly used observational study designs. In cohort studies, participants with an exposure of interest (e.g., a medication or lifestyle modification) are monitored over time for development of a particular outcome (the dependent variable such as heart attack, stroke, development of disease, side effect, etc.). In contrast, case control studies identify individuals with a particular outcome of interest and researchers retrospectively compare exposures between the two groups. However, the results of these studies should be interpreted with caution as these studies are subject to several biases.

### Cohort Studies

Cohort studies divide participants into groups based on whether they have experienced an exposure of interest.<sup>1</sup> Participants are followed over time to determine whether they develop the disease or outcome of interest. An example of an ongoing cohort study is the Millennium Family Cohort.<sup>2</sup> The study is evaluating the impact of military service on family

members with a 21 year follow up period. Families are grouped based on the deployment status of service members and outcomes of interest include the mental health, coping skills and well-being of military personnel and their family members.

Cohort studies are the best design for exploring potential relationships between rare exposures and development of an outcome (or disease), but are also widely used for common exposures as well. Cohort studies can be prospective or retrospective. Prospective studies follow participants from exposure until the outcome of interest occurs or the end of the observation period, while retrospective studies often utilize chart review from past patient records. Prospective studies are subject to less bias, but require

more time and resources than retrospective evaluations. The relationships between exposure and outcomes are often reported as a relative risk for experiencing the outcome between exposure groups (see part 2 of this series for a review of relative risk).<sup>1</sup>

Temporal effects (effects which may develop over time) can be evaluated in cohort studies as the exposures precede the outcome, which is one important aspect of determining causality. However, given the risk of confounding variables in cohort studies, results should be interpreted cautiously as there are many other aspects of causality which need to be considered.<sup>3</sup> Given that this temporal relationship exists and that research on risk factors through randomized control trials are often considered unethical (e.g., smoking),

TABLE 2. Comparison of Observational Study Designs

Study Type	Participant Groups (independent variable)	Study Outcome (dependent variable)	Prospective vs Retrospective
Scenario: A researcher is investigating the association between smoking and cancer			
Cohort Study	Participant groups based on whether patient has encountered exposure of interest (e.g., grouped as smokers and nonsmokers)	Compares the rates of development of cancer between the two groups (e.g., would determine the relative risk of developing cancer between the smoking and nonsmoking groups).	Prospective or retrospective
Case-Control Study	Participant groups based on whether or not patient has the outcome of interest (e.g., grouped based on having cancer or not having cancer)	Compares the risk of exposure in the outcome group compared to the control group (e.g., would determine the odds of having been exposed to smoking)	Retrospective
Nested Case-Control Study	Participant groups based on whether or not patient has the outcome of interest within a cohort study (e.g., grouped based on having cancer or not having cancer but all participants are taken from the same cohort)	Same as case-control study	Can be prospective within a cohort study

cohort studies are often utilized in this area.<sup>1</sup> An example of a large, cohort study now in its third generation of participants is the Framingham Heart Study in which researchers are looking to establish the effects of diet, exercise and medications on heart disease.<sup>4</sup>

### Case-Control Studies

In case-control studies, participants with an outcome of interest are retrospectively matched with control group participants who have not experienced the outcome.<sup>1</sup> Researchers retrospectively determine the risk of exposure for the participants in each group through the collection of past exposure data, and then evaluate how frequently they occur in each group. Choice of the control group is extremely important in case-control studies and can introduce bias if not chosen correctly. Everyone in the control group should have the opportunity to develop the outcome of interest; for example, men should not be included in a study of risk factors leading to ovarian cancer. Case-control studies can be used to help decide whether a specific exposure may have a relationship to the development of the outcome of interest, interest, or even to a rare side effect.<sup>1</sup> However, given their weaker study design they are often primarily considered to be hypothesis-generating studies. hypothesis-generating studies.

In some instances, case-control studies may be the only option for ethical reasons when studying rare outcomes or if there is a large time period between the exposures and the outcomes of interest. One example of such an occurrence involves the case-control study of the relationship between Creutzfeldt-Jakob disease and dietary risk factors.<sup>5</sup> This four year study separated participants based on whether or not Creutzfeldt-Jakob disease was diagnosed and utilized a participant survey to examine the consumption of various types of meats. This survey was used to verify the increased risk of Creutzfeldt-Jakob disease associated with the meat consumption.

Given that case-control studies are retrospective, these studies cannot be used to calculate relative risk directly; instead odd ratios are used to demonstrate relationships between the outcome and the exposure.<sup>6</sup> Relative risk cannot be directly calculated because these studies do not determine the risk of an outcome; rather the likelihood of

**TABLE 1. Bias in Observational Research**

<i>Type of Bias</i>	<i>Brief Description of Bias</i>	<i>Example situation when the bias could be present</i>
<b>Selection Bias</b>	The manner in which groups are determined results in study groups differing at baseline	A prospective cohort study is conducted investigating the relationship between ipratropium use and arrhythmias; however, patients who utilized ipratropium had more severe COPD.
<b>Information Bias</b>	Unequal collection of exposure data prompts a researcher to more preferentially search for a cause of the disease	A case-control study is investigating potential risk factors for a resistant infection. However, researchers collect exposure data face to face in a hospital for the infection group and over the phone for the control group.
<b>Recall Bias</b>	Patients with a disease have more “may be better able to remember past exposures a possible cause of the disease versus patients without the disease	A case-control study is investigating potential relationships between OTC medications and birth defects. Women who have children with birth defects may be more likely to remember medications that took than those with healthy children

being exposed is calculated.<sup>1</sup>

An odds ratio in a case-control study is interpreted differently from the traditional definition. An odds ratio from a randomized control trial determines the odds of developing an outcome amongst those in an exposure group compared to the odds in a control. In a case-control study, the odds ratio determines the odds of exposure amongst a group of participants with an outcome compared to the odds in a control without the outcome of interest. For example, an odds ratio from a hypothetical case-control study about deep vein thrombosis may find that a sedentary lifestyle is four times as frequent in participants who developed a thrombus compared to participants with an active lifestyle.

Although case-control studies are typically retrospective, an exception is the nested case-control design. A nested case-control is usually a sub-study “nested” within a cohort study.<sup>7</sup> Both outcome and control group participants are drawn from the original cohort study, and participants are prospectively followed from exposure to outcome. The benefit of conducting a nested case-control study is the minimization of recall bias (described in the following section) or errors in medical records data extraction.

### Biases

Selection bias occurs in an observational study when the two study groups differ

in some measured or unmeasured characteristics at baseline, or in the opportunity to develop the outcome being studied. Selection bias undermines the internal validity of an observational study, as it creates the question of whether the association found was truly due to what is being studied or due to a confounding variable such as differences between groups at baseline.

Information bias stems from inconsistent data collection between study groups.<sup>8</sup> For example in a case-control study, exposure information from those with the disease may be gathered bedside while a participant is hospitalized, whereas control group information may be gathered via telephone conversations. This difference in information gathering may trigger an observer to more thoroughly and preferentially research diseased participants for a cause. To prevent information bias, data collection in both cohort and case-control studies should be performed by a blinded observer who is unaware of the group allocations for each participant.

Another potential data collection bias is recall bias. Since retrospective studies utilizing interviews often rely on the memory of participants or family members to recall an exposure, those participants who have the outcome often have more incentive to try to recall more possibilities” to “tend to be more likely to recall potential exposures.<sup>8</sup> For example, a researcher might interview two groups of patients: those with

## New Members

Welcome to the newest members of PSW

Kristen Aschbrenner, Wausau  
Kenneth Baker, PBA Health, Kansas City, MO  
Timothy Battisti, Genoa City  
Heidi Nicole Belekevich, De Pere  
Victoria Brouner, Concordia University Wisconsin, Mequon  
William A Brown, Traverse City  
Rachel A Buchanan, Madison  
Evelyn J Buss, RPh, Poyette Hometown Pharmacy, Poyette  
Uvi Castillo, Concordia University Wisconsin, Mequon  
Amanda Coffey, Walgreens #11858, Madison  
Daniel M. Cunningham, RPh, ThedaCare Pharmacy, Appleton  
Mike Faber, Fond Du Lac  
Tammy Ferry, Concordia University Wisconsin, Mequon  
Kate Gainer, Iowa Pharmacy Association, Des Moines, IA  
Steven Galley, Walgreens Pharmacy, Ashland  
Cynthia Habeck, Tomah  
Holly Heckert, Milwaukee  
Julie Ann Hodgeman, Randolph  
Harry Horn, Middleton  
Susan Jaeck, Milwaukee  
Jennifer M Jones, CPhT, Appleton  
James Karnau, Ministry Saint Joseph's Hospital, Marshfield  
Mary D. Kraus, RPh, Waukesha  
Rebecca Lackey, Milwaukee  
Emily T Lin, Milwaukee  
Nicole Lubcke, Verona  
Amanda Ludwig, Madison  
Cassandra J Lytle, Merrimac  
Jennifer Marie Soto Meyer, Madison  
Nicole A Meyer, CPhT, Marshfield  
Jamie Moline, Greenleaf  
Paula Napier, PRh, Madison  
Emily Catherine Neumann, Somerset  
Roger E Overton, RPh, Onalaska  
Joe Revak, PharmD, Crystal Lake, IL  
Jeffrey T. Robertson, PharmD, Verona  
Ann E Schenkel, Palatine, IL  
Joyce Schubring, Aurora Pharmacy #1242, Marinette  
David Scofield, RPh, Hartig Drug Stores Corp, Dubuque, IA  
Kimberly Smithers  
Kathryn R Taylor, PharmD, Chippewa Falls  
Janelle Vittetoe, Madison  
Scott M Weber, RPh, Burlington  
Frances Wenzel, Kohler  
Kimberly White-Faull, Mineral Point  
Erin Wilkes, Milwaukee  
George Wilkinson, Concordia University Wisconsin, Mequon  
Matthew L Wolf, Pewaukee  
Michelle Yu, New Berlin

and without active ulcers. Patients with ulcers may be more likely to propose several contributing factors for their ulcer such as stress, NSAID use, or alcohol consumption. In contrast, patients without ulcers may not note these exposures as they did not experience the ulcers themselves. Both information and recall bias reduce internal validity as they call into question whether the observed differences between the two groups were due to differences in how the data were gathered.

Threats to external validity, can also be present in observational research and involve selection of the participant sample.<sup>8</sup> If inclusion criteria are too restrictive, the ability to generalize of the results to a larger population decreases. Other common biases that can occur in observational studies are attrition bias and potential confounding factors, which were defined in part 6 of this series.

### Conclusion

This article reviewed definitions and examples of case-control and cohort studies, and the different biases that can affect the internal and external validity of these observational studies.

### Practice Question

1. Observational studies are useful in which types of situations?
  - a. Unethical exposure risks
  - b. Rare disease states
  - c. Lengthy study time requirements
  - d. All of the above
2. Participants grouped by an outcome of interest and then have exposure risks retrospectively determined is an example of which type of study?
  - a. Cross-over Study
  - b. Cohort Study
  - c. Case-Control Study
  - d. Randomized Control Trial
3. A researcher who collects exposure data inconsistently between patient groups in a case-control study places the study at risk for which bias?
  - a. Information bias
  - b. Selection bias
  - c. Recall bias
  - d. A decrease in external validity

### Answers:

1. d. All of the reasons given are situations in

which observational studies are useful and other study designs may not be appropriate.

2. c. Case-control studies group each patient set together based on the absence or presence of the outcome of interest. Exposure risks for each group are then retrospectively determined to produce an odds ratio.
3. a. Information bias becomes a risk when a data collector in an observational study is prompted or gives preference to more thoroughly searching for exposure data based on inconsistent data collection settings.

---

Claire Seidler is a fourth year Doctor of Pharmacy student at the University of Wisconsin School of Pharmacy, Madison, WI. Amanda Margolis is a Lecturer at the University of Wisconsin School of Pharmacy and a Clinical Pharmacist at the William S. Middleton Memorial Veterans Hospital, Madison, WI. Kevin Look is an Assistant Professor in the Social and Administrative Sciences Division at the University of Wisconsin School of Pharmacy, Madison, WI.

*The authors of this article declare no real or potential conflicts or financial interest in any product or service mentioned in the manuscript, including grants, equipment, medications, employment, gifts and honoraria.*

### References

1. Mann CJ. Observational research methods. Research design II: cohort, cross sectional, and case-control studies. *Emerg Med J.* 2003;20(1):54-60.
2. Crum-Cianflone NF, Fairbank JA, Marmar CR, Schlenger W. The Millennium Cohort Family Study: a prospective evaluation of the health and well-being of military service members and their families. *Int J Methods Psychiatr Res.* 2014. [Epub ahead of print]
3. Steiner PM, Cook TD, Shadish WR, Clark MH. The importance of covariate selection in controlling for selection bias in observational studies. *Psychol Methods.* 2010;15(3):250-267.
4. Mahmood SS, Levy D, Vasan RS, Wang TJ. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. *Lancet.* 2014;383(9921):999-1008.
5. Davanipour Z, Sobel E, Ziogas A, et al. Dietary Risk Factors for Sporadic Creutzfeldt-Jakob Disease: A Confirmatory Case-Control Study. *Br J Med Med Res.* 2014;4(12):2388-2417.
6. Sedgwick P. Case-control studies: measures of risk. *BMJ.* 2013;346:f1185.
7. Ernster VL. Nested case-control studies. *Prev Med.* 1994;23(5):587-590.
8. Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet.* 2002;359(9302):248-252.