



# HHS Public Access

Author manuscript

*Am J Nurs.* Author manuscript; available in PMC 2022 October 06.

Published in final edited form as:

*Am J Nurs.* 2021 October 01; 121(10): 59–62. doi:10.1097/01.NAJ.0000794280.73744.fe.

## Overview: Cross-Sectional Studies

**Bernadette Capili, PhD, NP-C [Director]**

Heilbrun Family Center for Research Nursing, The Rockefeller University, 1230 York Avenue, Hospital, Room 106, New York, NY 10065

---

The conduct of research requires the selection of the appropriate method to evaluate the research problem or question. Due to some topics' ethical nature or the need to understand the natural history (i.e., disease or condition), using an observational study design might be the best fit. The primary purposes of observational studies are to describe and examine the distributions of independent (predictor) and dependent (outcome) variables in a population (sample) and analyze the associations between them (Cummings, 2013). Observational studies monitor study participants without providing study interventions. This paper describes the cross-sectional design, examines the strengths and weaknesses, and discusses some methods to report the results. Future articles will focus on other observational methods, the cohort, and case-control designs.

### Cross-Sectional Design

Cross-sectional designs help determine the *prevalence* of a disease, phenomena, or opinion in a population, as represented by a study sample. *Prevalence* is the proportion of people in a population (sample) who have an attribute or condition at a specific time point (Mann, 2012) regardless of when the attribute or condition first developed (Wang & Cheng, 2020). Additionally, each study participant's evaluation is completed at one time-point with no follow-ups (Cummings, 2013), providing a 'snapshot' of the sample. Cross-sectional designs can be implemented as an interview or survey and may also collect physiological data and biological samples.

### Cross-Sectional Design: Descriptive

Cross-sectional studies can be descriptive and analytic (Alexander, 2015a). Descriptive cross-sectional studies characterize the *prevalence* of health outcomes or phenomena under investigation. *Prevalence* is measured either at a one-time point (*point prevalence*), over a specified period (*period prevalence*) (Alexander, 2015a), or as a cross-sectional serial survey (Cummings, 2013). The descriptive design starts by identifying the population of interest, collects the data, and classifies the participant, either as having the outcome or phenomena of interest or not (Mann, 2012). For example, investigators want to determine the *point prevalence* of obesity among people with HIV. To conduct this study, investigators select several HIV primary care clinics in their region and obtain heights, weights, and measure waist circumference during one specified day at each clinic. For a *period prevalence* study,

the investigators could visit each clinic at four-time points over 12 months to obtain body measurements to capture other patients visiting the clinics. *Period prevalence* and *point prevalence* are similar, except that the time-frame is broader since it can be difficult to evaluate or observe the entire population or sample at one time-point\*\*.

For serial cross-sectional surveys, investigators collect data in the same population over a specified period. It uses a longitudinal time-frame. For example, every three years, investigators repeat the body measurements among HIV patients to draw inferences about the patterns over time about obesity (Cummings, 2013). However, new samples are selected each time; therefore, each participant's changes cannot be evaluated. It is important to note that the results may be affected by "people entering or leaving the population due to births, deaths, and migration" (Cummings, 2013, p.88).

### Method to Report Results: Descriptive Cross-Sectional Design

*Prevalence* is generally reported as a percentage (30% or 75 out of 250 HIV patients were obese). Knowing the *prevalence* of a condition in a population (sample) helps understand the disease burden in terms of services needed, morbidity, mortality, and quality of life (Noordzij, Dekker, Zoccali, & Jager, 2010). For instance, if obesity is high among the participants, clinic visits could provide nutritional counseling and physical activity recommendations and regularly monitor body weight measurements to prevent the complications associated with obesity (i.e., knee osteoarthritis, type 2 diabetes mellitus).

$$Prevalence = \frac{\text{Number of participants with the condition at the time point}}{\text{Total number of participants in the sample}}$$

### Cross-Sectional Design: Analytic

Analytic cross-sectional studies can provide the groundwork to infer preliminary evidence for a causal relationship (Mann, 2012). This design allows investigators to identify a population or sample and collect *prevalence* data to evaluate outcome differences between exposed and unexposed participants on a disease, phenomena, or opinion (Wang & Cheng, 2020). This design compares the proportion of participants exposed to the disease or phenomena of interest with the proportion of participants non-exposed with the disease or phenomena of interest (Alexander, 2015a). However, determining which variable is the dependent and independent variable or cause and effect is difficult to determine. For example, the association between obesity and hours spent in sedentary behavior among HIV patients (see Table 1). Which came first? Did the participant become obese due to sedentary behavior, or was the participant inactive due to obesity? According to Cummings et al., 2013, determining which variable to label as dependent or independent "depends on the cause-and-effect hypotheses of the investigator" (p. 85) or the biological plausibility rather than on the study design.

### Method to Report Results: Analytic Cross-Sectional Design

In continuing with the obesity and sedentary activity level among HIV participants, the example below (see Table 1) describes the methods for calculating and discussing the results

for an analytic cross-sectional study. The *prevalence odds ratio (POR)* (calculated as  $[ad/bc]$ ) and *prevalence ratio (PR)* (calculated as  $[a/(a + b)] / [c/(c + d)]$ ) are commonly used to report estimates of association between independent and dependent variables in cross-sectional studies (Tamhane, Westfall, Burkholder, & Cutter, 2016).

### Prevalence Odds Ratio/Odds Ratio

The *POR* is calculated similarly to the *odds ratio (OR)* (Alexander, 2015b) and referred to as *POR* when *prevalence* is used (Tamhane et al., 2016). *OR* measures the association between exposure and outcome (see Table 1) and denotes the chances that an outcome happens with a specific exposure, compared to the chances of an outcome happening in the absence of the exposure (Szumilas, 2010). This information helps both clinicians and investigators determine if certain factors (i.e., clinical characteristics, medical history) are a risk for a particular outcome (i.e., disease, condition). Future studies or health policies can target methods to prevent or treat outcomes (i.e., disease, condition) identified in such studies.

For example, in Table 1, using the formula and dataset below, the *OR* was 2.4. The result shows that the obese HIV participants (exposed) were two and a half times (2.5x) more likely to be sedentary than the non-obese participants (unexposed). If the *OR* for the dataset was equal to 1, then the exposure (obese) did not affect the outcome's odds. In other words, the chance of being sedentary is the same in the exposed (obese) and the non-exposed (not obese) groups. Similarly, if the *OR* was less than 1, it implies that the exposed (obese) group, were less likely to be sedentary (outcome) compared to the non-obese group (unexposed) (Tenny & Hoffman, 2019).

### Prevalence Ratio/Risk Ratio and Excess Prevalence/Risk Difference

The *PR* is calculated similarly to the *risk ratio (RR)* (Alexander, 2015b). The *PR* measures the *prevalence* of an outcome in the exposed group, divided by the unexposed group, and measures the association's strength between the exposure and outcome (Alexander, 2015). *Excess prevalence (EP)* or the *risk difference (RD)* provides the difference in *prevalence* between the groups and indicates how much additional *prevalence* is due to the exposure of interest (Alexander, 2015b). From Table 1, the *PR/RR* for the example equaled 2.07, with an *EP* of 11.9%. The results might conclude that obesity among the HIV participants was twice (2.07) as common and occurred almost 12% more often among HIV participants who were sedentary.

Similar to the *OR* interpretation, if the *RR* was equal to 1, exposure did not prevent or harm the exposed and unexposed groups. In other words, being obese did not affect the activity level (sedentary versus not sedentary). If the *RR* was less than 1, it implies that the exposure had a protective effect in that obese HIV participants were less likely to be sedentary than the unexposed group (not obese).

### Considerations for use: Prevalence Odds Ratio versus Prevalence Ratio

The statistical literature has numerous articles discussing the pros and cons of using either the *POR/OR* or *PR/RR* for cross-sectional studies (Tamhane et al., 2016). Consulting a

statistician to discuss the best choice for each project is highly recommended. However, according to Alexander and colleagues (2015a), the *POR* is preferred when the study topic is a chronic condition (i.e., hypertension, HIV), or the risk of developing the disease takes several months to develop. For studies evaluating acute conditions (i.e., the common cold), the *PR* is favored (Alexander, 2015a).

Furthermore, suppose the *prevalence* of a disease or phenomena is low, less than ten percent in the exposed and unexposed population (sample). In that case, the resulting *POR* and *PR* will be equal (Alexander, 2015a). Since cross-sectional studies are suitable for examining chronic diseases or conditions, the *POR* is generally the ideal measure of association to use (Alexander, 2015a).

## Confidence Intervals

*Confidence intervals* (CI) measure the precision of the OR, RR, or the possible “variation in a point estimate (the mean value)” (Alexander, 2015b, p 4). A narrower *CI* indicates a higher level of precision versus a wider *CI* suggesting a lower level of precision (Cummings, 2013). The sample size also impacts the *CI*’s width, with larger sample sizes providing a more precise estimate. The approximate value of the point estimate is based on factors (i.e., characteristics like body weight, level of activity) such as the mean (average) of a population from a population’s random samples.

From Table 1, the *OR* = 2.4 with a confidence interval of (95% *CI* (1.4713 – 3.9150)) might conclude that the obese HIV participants were two and a half times (2.5x) more likely to be sedentary than the non-obese participants. 2.4 is the point estimate obtained from this example; however, the entire population of obese HIV people was not included. If other samples of HIV participants were assessed, the point estimate would likely differ. Some samples might get the point estimate of less than or some greater than 2.4.

The 95% *CI* is the interval representing the (population) parameter value 95% of the time if an experiment or study is repeated, in that 95 out of 100 intervals would result in the intervals containing the true risk ratio or odds ratio value. For the sedentary and obesity study, the interpretation might conclude that a 2.4 point estimate could range from a low of 1.4713 to a high of 3.9150.

## Strengths

The main strength of the cross-sectional design is the ability to obtain results faster. Investigators do not need to wait for outcomes to occur. Participants either have the condition or attribute at the time of data collection or not. Furthermore, there are no participant follow-ups; therefore, losing study participants during the study is not an issue.

The design’s inherent nature makes it inexpensive to conduct and can yield multiple independent (predictor) and dependent (outcome) variables (Cummings, 2013). The data collected can lead to additional studies to build upon the knowledge obtained. From the example, the investigators learned that obese HIV participants were more likely to be

sedentary; the next study might develop a clinical trial to determine the methods to increase activity level in this population.

## Weaknesses

A significant limitation of using this design is the inability to measure the *incidence* of a disease or attribute (Wang & Cheng, 2020). *Incidence* measures the proportion of participants that develop a disease or attribute over time (Cummings, 2013). In other words, investigators need a follow-up phase to determine the *incidence*. In continuing with the example, if investigators continued to follow the HIV participants who were obese but not sedentary, would additional time (follow-up) result in increased sedentary behavior associated with conditions secondary to aging or worsening of immune status? Unfortunately, the cross-sectional design can not answer this question.

*Additionally, the prevalence* of a disease or attribute is influenced by the *disease's incidence and survival or disease duration* (Alexander, 2015a). For example, participants who live longer with a disease will have a higher likelihood of being counted ( $Prevalence = \# \text{ of participants with the condition at the time point} / \text{Total \# of participants in the sample}$ ) versus those who are short-term survivors. Moreover, if treatments for a disease or attribute are improved, or the survival time-frame decreases, the disease or attribute's prevalence will reduce (Alexander, 2015a). New information presented to the lay public could also influence the *prevalence* of a disease or attribute through lifestyle changes (i.e., increasing physical activity, improving diet) or changing jobs if the profession is associated with an identified risk or disease. Therefore, this design does not allow investigators to ascertain the events' sequence, which came first, obesity or sedentary behavior.

For investigators studying rare diseases or conditions, the cross-sectional design is not the best fit. Cross-sectional studies often draw samples from a large and heterogeneous study population (Wang & Cheng, 2020). Participants with the rare condition of interest might not be identified in the study sample.

## Reporting Recommendations

A reporting guideline for cross-sectional studies is available for investigators and consumers of research to use. A reporting guideline's primary goal is to ensure that published clinical research studies provide transparency in reporting a study's conduct (what was done) and results. The guideline is a tool investigators can use to develop their manuscripts and offers a checklist of inclusion items for a published paper (Equator.network). The recommended items will help ensure that a reader can understand the manuscript, follow the study's planning and how the research was conducted, the findings, and the conclusions (von Elm et al., 2014).

For cross-sectional studies, the guideline is titled *Strengthening the Reporting of Observational Studies in Epidemiology* (STROBE) (von Elm et al., 2014). The *STROBE* guideline is a 22-item checklist. The checklist provides essential information for a study to be replicated, useful for healthcare professionals to make clinical decisions, and give

enough information for inclusion in a systematic review (<https://www.equator-network.org/reporting-guidelines/strobe/>).

## Conclusion

The cross-sectional design is an appropriate method to determine the prevalence of a disease, attribute, or phenomena in a study sample. The design provides a ‘snapshot’ of the sample, and investigators can describe their study sample and review associations between the collected variables (independent and dependent). The observational nature makes it relatively quick to complete a study and provides data to support future studies that might lead to methods to treat or prevent diseases or conditions.

## Acknowledgments

This manuscript is supported in part by grant # UL1TR001866 from the National Center for Advancing Translational Sciences (NCATS), National Institutes of Health (NIH) Clinical and Translational Science Award (CTSA) program.

## References

- Alexander L, Lopes B, Richetti-Masterson K, Yeatts KR. (2015a) Cross-Sectional Studies. In: Vol. Second Edition. ERIC Notebook. Durham, NC: Department of Epidemiology at the UNC Gillings School of Global Public Health.
- Alexander L, Lopes B, Richetti-Masterson K, Yeatts KR. (2015b) Risk and Rate Measures in Cohort Studies. In: Vol. 2nd. ERIC Notebook. Durham, NC: Department of Epidemiology at the UNC Gillings School of Global Public Health.
- Cummings S B. H. S. R. (Ed.) (2013). *Designing Clinical Research* (4th ed.). Philadelphia, PA: Wolters Kluwer/Lippincott Williams & Wilkins.
- Mann CJ (2012). Observational research methods—Cohort studies, cross-sectional studies, and case-control studies. *African Journal of Emergency Medicine*, 2(1), 38–46. doi:10.1016/j.afjem.2011.12.004
- Noordzij M, Dekker FW, Zoccali C, & Jager KJ (2010). Measures of disease frequency: prevalence and incidence. *Nephron Clin Pract*, 115(1), c17–20. doi:10.1159/000286345 [PubMed: 20173345]
- Tamhane AR, Westfall AO, Burkholder GA, & Cutter GR (2016). Prevalence odds ratio versus prevalence ratio: choice comes with consequences. *Stat Med*, 35(30), 5730–5735. doi:10.1002/sim.7059 [PubMed: 27460748]
- Tenny S, & Hoffman MR (2019). *Odds Ratio (OR)*: StatPearls Publishing, Treasure Island (FL).
- von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, & Vandenbroucke JP (2014). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *Int J Surg*, 12(12), 1495–1499. doi:10.1016/j.ijsu.2014.07.013 [PubMed: 25046131]
- Wang X, & Cheng Z (2020). Cross-Sectional Studies: Strengths, Weaknesses, and Recommendations. *Chest*, 158(1), S65–S71. doi:10.1016/j.chest.2020.03.012 [PubMed: 32658654]

**Table 1:**

Calculation Example

Outcome	<u>Exposed</u> <i>Obese</i> (Body Mass Index ≥ 30)	<u>Unexposed</u> <i>Not Obese</i> (Body Mass Index < 30)	Total
<u>Disease</u> <i>Sedentary</i> *(Low Activity Level)	75 <i>a</i>	250 <i>b</i>	325 (a + b)
<u>No Disease</u> <i>Not Sedentary</i> *(Moderate to High Activity Level)	25 <i>c</i>	200 <i>d</i>	225 (c + d)
<b>Total</b>	100 (a + c)	450 (b + d)	550 (a + b + c + d)

\* measured by the Physical Activity Questionnaire

- a = exposed participant and acquires the outcome of interest
- b = unexposed participant and acquires the outcome of interest
- c = exposed participant and does not acquire the outcome of interest
- d = unexposed participant and does not acquire the outcome of interest

$$Prevalence = a + c / a + b + c + d = 100 / 550 = .1818 \times 100 = 18.18\%$$

Prevalence of HIV participants who are obese and sedentary =  $a / (a + b) = 75 / 325 = .23 \times 100 = 23\%$

Prevalence of HIV participants who are obese and not sedentary =  $c / (c + d) = 25 / 225 = .11 \times 100 = 11.1\%$

Prevalence of overall HIV participants who are obese =  $(a + c) / (a + b + c + d) = 100 / 550 = .182 \times 100 = 18.2\%$

$$Prevalence\ odds\ ratio / Odds\ Ratio = ad / bc = 75 \times 200 / 250 \times 25 = 15000 / 6250 = 2.4$$

Interpretation of Prevalence Odds Ratio/Odds Ratio:

- OR = 1 Exposure did not effect the odds of the outcome
- OR > 1 Exposure is associated with the higher odds of outcome versus nonexposed group
- OR < 1 Exposure is associated with lower odds of outcome versus exposed group

\*\* Confidence Intervals results for the OR = 95% CI (1.4713 – 3.9150)

- Upper 95% CI =  $e^{[\ln(OR) + 1.96 \sqrt{(1/a + 1/b + 1/c + 1/d)}]} = 1.4713$

- Lower 95% CI =  $e^{[\ln(OR) - 1.96 \sqrt{(1/a + 1/b + 1/c + 1/d)}]} = 3.9150$

$$Prevalence\ Ratio / Risk\ Ratio = \frac{a / (a+b)}{c / (c+d)} = \frac{23\%}{11.1\%} = 2.07$$

$$\text{Excess Prevalence/Risk Difference} = a/(a + b) - c/(c + d) = 23\% - 11.1\% = 11.9\%$$

Interpretation of Prevalence Ratio/Risk Ratio:

RR = 1 Exposure did not prevent or harm the exposed and unexposed groups

RR > 1 Exposure is harmful to the exposed group compared to the unexposed group

RR < 1 Exposure is less harmful (protective) to the exposed group compared to the unexposed group

\*\*\* Confidence Intervals results for the RR = 95% CI (1.3653 – 3.1595)

- Upper 95% CI =  $e^{\ln(\text{RR}) - 1.96 \sqrt{1/a + 1/c - 1/a + b - 1/c + d}}$  = 1.3653

- Lower 95% CI =  $e^{\ln(\text{RR}) + 1.96 \sqrt{1/a + 1/c - 1/a + b - 1/c + d}}$  = 3.159

References: Alexander, 2015a, Cummings, 2013, Tenny & Hoffman, 2019.

\*\*[https://www.medcalc.org/calc/odds\\_ratio.php](https://www.medcalc.org/calc/odds_ratio.php) (web-based confidence interval calculator of odds ratio)

\*\*\*[https://www.medcalc.org/calc/relative\\_risk.php](https://www.medcalc.org/calc/relative_risk.php) (web-based confidence interval calculator RR)