International Journal of
Mathematics And its Applications
ISSN: 2347-1557

Int. J. Math. And Appl., **13(1)**(2025), 37–45
Available Online: http://ijmaa.in

Cross-sectional Studies in Clinical Researches

Randa Eltayeb[1,3,*], Adil Younis[1,2], Mohammed Elamin[2], F. A. Al Mostafa[1]

[1]*Department of Mathematics, College of Business & Economics, Qassim University, Buraydah, Saudi Arabia*
[2]*Department of Statistics, Faculty of Science, Sudan University of Science and Technology, Khartoum, Sudan*
[3]*Department of Mathematics, Sennar University, Sennar, Sudan*

Abstract

Cross-sectional studies are observational studies in which the investigator measures the exposure and outcome at a single point of time. Observational studies means that, the investigator only observes and makes no interventions to the participants. Cross-sectional studies also known as prevalence studies are used frequently in the developing countries since they are relatively inexpensive and easy to conduct compared to other types of studies. This paper reviews the definition, types, advantages and disadvantages of the cross-sectional studies.

Keywords: Cross-sectional studies; bias; prevalence; exposure; prevalence ratio.

1. Introduction

Cross-sectional studies measure outcome and exposure simultaneously in a given population. They provide a "snapshot" of the outcome and exposure characteristics in a population or representative sample at a specific point in time with no follow-up period. Cross-sectional studies serve many purposes, and the cross-sectional design is the most relevant design when assessing the prevalence of disease, attitudes and knowledge among patients and health personnel [4]. It usually mentioned that data in cross sectional research is collected at a given point in time, but what is meant by it is not explained nor defined. The investigator draws a representative sample from the population and classifies the individuals according to output and exposure status see figure 1. A cross-sectional design is relevant when assessing the prevalence of disease or traits, attitudes and knowledge, in validation and in reliability studies. When cross-sectional data is used for analytical purposes, authors and readers should be careful not to make causal inferences, unless the exposure may safely be assumed to be stable over time [4]. Sometime, determining the output and exposure based on biologic plausibility and historical information [2]. For example Scholten Mia and et. al. (2022), were interested in studying disparities in prevalence of heart failure according to age, multi-morbidity level and socioeconomic

*Corresponding author (fazasluman@gmail.com)

status in southern Sweden, they found that the prevalence of HF increased with advancing age and the multi-morbidity level also HF had a strong correlation with the socioeconomic status of the primary healthcare centers with the most significant disparity between 40 and 80 years of age. Obviously the investigators dealt with age, multi-morbidity level and socioeconomic status variables as exposures and heart failure as outcome.

2. Types of Cross-sectional Studies

Cross-sectional studies can be classified either descriptive or analytical, depending on the objective of the study. Descriptive cross- sectional studies simply calculate the prevalence of diseases or exposures in a specific population [3].

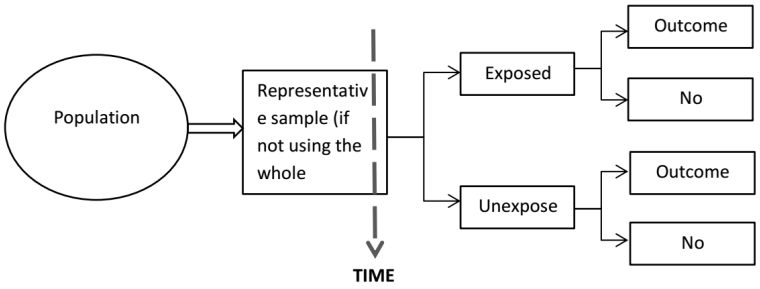

Figure 1: Design of cross-sectional studies

In analytical cross-sectional studies, the investigator examines the association between outcome and exposure but they do not allow inferences related to temporal sequences between exposure and disease [5] except for constitutional factors such as age, race, and sex; these cannot be altered by other variables and therefore are always predictors [2]. Analytical cross-sectional studies are useful for hypothesis generation for potential association between the exposure and the disease that can be confirmed or rebutted using more trusted study designs e.g. a cohort or clinical trial studies [6]. There is a subtype of cross sectional study called repeated (or serial) cross-sectional study in which, the same (or similar) information is asked to a different sample of individuals each time" [15] on the same target population. Repeated cross-sectional studies can help to answer and track "how many" or to what extent, but it is not possible to answer "why" change took place [7]. Repeated cross sectional studies have a longitudinal time frame. They look for population changes over time (also known as aggregate change over time), so changes within individuals cannot be assessed as in a cohort studies because a new sample is drawn each time [2, 3].

3. Sampling in Cross Sectional Studies

Planning the sampling strategy is an essential component of cross-sectional study design. In epidemiology, sampling can be defined as the process of selecting certain members or a subset of the whole population to estimate the characteristics of the population. Creating a solid sampling plan in

a cross sectional study is critical because of the considerable heterogeneity usually observed in the target population [3]. Table 1 shows different type of sampling techniques.

4. Bias

Investigators should be aware of bias when planning a cross-sectional study. Bias may be defined as any systematic error in a study that results in an incorrect estimate of the true effect of an exposure on the outcome of interest. There are many types of bias in clinical studies, but for simplicity, they can be broadly grouped into two categories: selection bias and information bias [3].

Probability sampling methods	
Simple random sampling	Every member of the population has the same probability of being randomly selected into the sample
Systematic sampling	One selects every n^{th} (ie, 8^{th}) subject in the population to be in the sample
Stratified sampling	The population is divided into non-overlapping groups, or strata; a random sample of population members is then collected from within each stratum
Clustered sampling	The researcher divides the population into separate groups, called clusters. Then, a simple random sample of clusters is selected from the population. Note that the clusters are used as the sampling unit, rather than individuals
Nonprobability sampling methods	
Convenience sampling	Participants are selected based on availability and willingness to take part
Quota sampling	A tailored sample that is in proportion to some characteristic or trait of a population
Purposive sampling	Also known as judgmental or subjective sampling. It relies on the judgment of the researcher when choosing members of the population to participate in a study
Snowball sampling	Existing study subjects recruit future subjects from among their acquaintances

Table 1: Some Sampling Methods used in Clinical Studies

Selection biases are distortions that result from procedures used to select subjects and from factors that influence study participation. The common element of such biases is that the relation between exposure and disease is different for those who participate and for all those who should have been theoretically eligible for study, including those who do not participate [1]. A common type of selection bias is the nonresponse bias, which is usually encountered in cross-sectional survey studies with mailed questionnaires. A nonresponse bias occurs when the characteristics of non-responders differ from responders. Prevalence-incidence bias (also called the Neyman bias) is also particularly common in cross-sectional studies. It is a type of selection bias that occurs when the selection process favors

individuals with characteristics that are not representative of the population as a whole. For example, if the inclusion/exclusion criteria or sampling method leads to fewer subjects with mild disease in a study, an error in the estimated association between an exposure and an outcome could be seen [3]. An information bias occurs during data collection. The most important type of information bias is the misclassification bias. A misclassification bias is present when the detection of the exposure status (exposure identification bias) and/or the disease assessment (disease identification bias) is biased, i.e. exposed/diseased individuals are classified as non-exposed/non-diseased and vice versa. Misclassification can be non-differential or differential [8]. Recall bias and detection bias are two common information biases. Because exposure and outcome are measured simultaneously in a cross-sectional study, prior knowledge of the condition might influence the ascertainment of the exposure or the outcome, which results in recall bias [3].

Confounding: Confounding is a distortion in an observed relationship between exposure and outcome that is brought about by a third factor (the confounding factor or confounder) that is associated with both the outcome of interest and the exposure. In other words, confounding is an alternative explanation for an observed relationship [5].

For a variable to be a confounder, it should meet three conditions. The variable must: (1) be associated with the exposure being investigated; (2) be associated with the outcome being investigated; and (3) not be in the causal pathway between exposure and outcome. Confounding could result in a distortion of the association between exposure and outcome. Many statistical techniques may be applied to prevent or control for confounding. These include restriction, stratification, and matching [3].

Measuring Data: The prevalence ratio and prevalence odds ratio are two ways to present the association in the cross-sectional studies. PORs, often referred to as prevalence ORs when comparing prevalence can be calculated in which exposures, as well as outcomes (eg the prevalence of a particular disease), are studied at a similar point in time. Similarly, prevalence ratios can be calculated in place of relative risks (eg in cohort studies). Prevalence ratios are calculated in the same manner as relative risks, but have a different interpretation [9].

The Prevalence Ratio: The Prevalence Ratio (PR), which is defined in terms of "how many times are exposed individuals more likely to have the disease or condition than unexposed individuals [1, 6]. PR is estimated as the prevalence of the disease in exposed group divided by the prevalence of the disease in unexposed group.

In the following, we will illustrate how to calculate the prevalence ratio from a 2×2 contingency table.

Exposure	Disease		Total
	Yes	No	
Yes	a	b	$a+b$
No	c	d	$c+d$
Total	$a+c$	$b+d$	n

Table 2: 2×2 contingency table for disease and exposure

The prevalence of the disease in the exposed group, based on table 1 will be calculated as follows:

$$P_{exp} = \frac{a}{a+b}$$

The prevalence of the disease in the unexposed group, based on table 1 will be calculated as follows:

$$P_{un-exp} = \frac{c}{c+d}$$

Finally, the PR will be calculated from the ratio of the prevalence in exposed group to the prevalence in unexposed group as follow:

$$PR = \frac{P_{exp}}{P_{un-exp}} = \frac{a/(a+b)}{c/(c+d)}$$

The value of PR can be any number in the interval $(0, \infty)$. An PR of 1.00 means that the prevalence of disease is identical in the exposed and unexposed. An PR that is less than 1.00 means that the prevalence is lower in the exposed group. An PR that is greater than 1.00 means that the prevalence is increased in the exposed group. Paddy Ssentongo and et. al. (2020) conducted a cross-sectional study to assess the association of vitamin A deficiency (VAD) with early childhood stunting in Uganda.
The following table presents some data obtained in the survey:

Vitamin A deficiency status	Undernutrition		Total
	Yes	No	
Yes	341	83	424
No	2753	1588	4341
Total	3094	1671	4765

Table 3: Association between exposure/risk factor (Vitamin A deficiency) and disease (undernutrition) in children aged 6-59 months, Demographic and Health Surveys conducted in Uganda [14]

From table 3, the prevalence of the undernutrition with VAD is:

$$P_{exp} = \frac{341}{424} = 0.804$$

and the prevalence of the undernutrition without VAD is:

$$P_{un-exp} = \frac{2753}{4341} = 0.634$$

Therefore the prevalence ratio is

$$PR = \frac{0.804}{0.634} = 1.27$$

The interpretation of PR can be stated in several ways, e.g.:

- The prevalence of the undernutrition with VAD is 1.27 times that without VAD.
- The prevalence of the undernutrition with VAD is 127% that without VAD.

- VAD is associated with a 0.27 (1.27-1) fold increase in the prevalence of undernutrition.
- VAD is associated with a 27% increase in the prevalence of undernutrition.

The PR can be obtained from binomial regression or poisson regression models.

Prevalence odds ratio: The POR is the odds of an event in the exposed group divided by the odds of that event in the unexposed group. Returning to table 1, the odds of the disease in the exposed group will be calculated as follows:

$$Odd_{exp} = \frac{a}{b}$$

The odds of the disease in the unexposed group will be calculated as follows:

$$Odd_{un-exp} = \frac{c}{d}$$

Finally, the POR will be calculated from the ratio of the odds of the disease in the exposed group to odds of the disease in the unexposed group as follow:

$$POR = \frac{Odd_{exp}}{Odd_{un-exp}} = \frac{a/b}{c/d} = \frac{a.d}{b.c}$$

Like Prevalence ratio, The POR can be any number in the interval $(0, \infty)$. An POR of 1.00 means that the prevalence of disease is identical in the exposed and unexposed. An POR that is less than 1.00 means that the prevalence is lower in the exposed group. An POR that is greater than 1.00 means that the prevalence is increased in the exposed group. With the reference to the data in table 3, the odds of developing undernutrition with VAD are 341:83 (i.e, 341/32 or 10.66), and the odds of developing undernutrition without VAD are 2753:1588 (i.e, 2753/1588 or 1.73). Therefore, the POR is 10.66/1.73, or 6.16. The interpretation of POR can be stated in several ways, e.g.:

- The odds of developing undernutrition with VAD relative to without VAD are 6.2 to 1.
- The odds of developing undernutrition with VAD are about 6.2 as large compared without VAD .
- The odds of developing undernutrition with VAD are 502% higher compared without VAD.

The POR can be obtained from Logistic regression model, which is available in many Software statistical packages. The interpretation of POR is more confusion compared to the prevalence ratio. Prevalence odds ratios are routinely misinterpreted as prevalence ratios, the two measures present the association between the exposure and the outcome but the POR is always overestimated to the PR, look to the values of PR and POR from table 2, ($PR = 1.27$, $POR = 6.2$); these two values are approximately the same when the prevalence of disease is low, say less than 10 [13].

Advantages of cross-sectional studies:

- Less expensive and time consuming compared to other types of studies.
- Suitable for studying multiple outcomes and exposures at the same time.
- There is no loss to follow-up.
- Useful for public health planning, understanding disease etiology and for the generation of hypotheses [10].

Disadvantages of cross-sectional studies:

- Because the exposure and outcome are simultaneously assessed, there is no evidence of a temporal relationship between the exposure and the outcome [11] (like egg and hen problem, which one come first??). That is the investigator may conclude that there is an association (positive or negative) between an exposure and an outcome but there is no evidence that the exposure caused the outcome [12].
- Not good for rare diseases or for diseases of short duration due to low prevalence at a single point in time [11].
- Susceptible to biases such as nonresponse bias and recall bias
- Cross-sectional studies evaluate prevalent rather than incident outcomes [12].So they susceptible to prevalence-incidence bias (Neyman bias).

Cross-sectional studies have a problem, often called length-biased sampling is that the cases identified in a cross-sectional study will over represent cases with long duration and under represent those with short duration of illness [1].

5. Conclusion

Cross-sectional studies are observational studies which can be classified into descriptive or analytical studies based on the objective of the study. Descriptive studies are useful for assessing the prevalence of diseases, whereas analytical studies useful for assessing the association and for hypothesis generation. In analytical cross-sectional studies, the association can be measured using the prevalence ratio or prevalence odds ratio, the investigator must be aware to choose the suitable measure and report it correctly. The association that the study shows may be susceptible to many types of bias such as nonresponse, recall and prevalence-incidence bias. When reviewing a cross-sectional study, a reviewer should consider commenting on the following:

- The study population and the research question. Was the study population appropriate for the research question? Were there potential sources of bias related to the methods used to sample the population of interest? How was the possibility of selection bias addressed in the study design or analysis?

- The exposure(s), outcome(s), and relevant covariates. Are they clearly defined? Are there potential biases related to the accuracy of their measurement or the techniques used to collect data? How were missing data managed? The analysis and interpretation of the findings. Were potential confounders identified? Were potential confounders managed appropriately in the study design and/or analysis? If a regression model was built, were variables selected appropriately? Given the observational study design and strength of the association(s) identified, were the findings properly interpreted [3]?

References

[1] K. J. Rothman, S. Greenland and T. L. Lash, *Modern Epidemiology*, Vol. 3, Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins, (2008).

[2] S. B. Hulley, S. R. Cummings, W. S. Browner, D. G. Grady and T. B. Newman, *Designing clinical research*, 4th Edition, Lippincott Williams and Wilkins, (2013).

[3] X. Wang and Z. Cheng, *Cross-Sectional Studies: Strengths, Weaknesses and Recommendations*, Chest Journal, 158(2020), S65-S71.

[4] U. S. Kesmodel, *Cross-sectional studies - what are they good for?*, Acta Obstet Gynecol Scand., 97(4)(2018), 388-393.

[5] A. Chattopadhyay, *Oral health epidemiology: Principles and practice*, Jones and Bartlett Publishers, (2011).

[6] N. Pandis, *Cross-sectional studies*, American Journal of Orthodontics and Dentofacial Orthopedics, 146(1)(2014), 127-129.

[7] T. Doering, N. C. Suresh and D. Krumwiede, *Measuring the effects of time: repeated cross-sectional research in operations and supply chain management*, Supply Chain Management, 25(1)(2020), 122-138.

[8] G. Tripepi, K. J. Jager, F. W. Dekker and C. Zoccali, *Selection bias and information bias in clinical research*, Nephron Clin Pract., 115(2)(2010), c94-c99.

[9] D. P. Mc Kenzie and C. Thomas, *Relative risks and odds ratios: Simple rules on when and how to use them*, European Journal of Clinical Investigation, (2020), e13249

[10] K. A. Levin, *Study design III: Cross-sectional studies*, Evidence-Based Dentistry, 7(1)(2006), 24-25.

[11] S. R. Taur, *Observational designs for real-world evidence studies*, Perspectives in clinical research, 13(1)(2022), 12-16.

[12] M. D. Carlson and R. S. Morrison, *Study design, precision, and validity in observational studies*, Journal of Palliative Medicine, 12(1)(2009), 77-82.

[13] A. Indrayan, *Relative risk, odds ratio, attributable risk and number needed to treat*, Basic Methods of Medical Research, Third Edition, (2012).

[14] P. Ssentongo, Djibril M Ba, A. E. Ssentongo, C. Fronterre, Andrew Whalen, Yanxu Yang, J. E. Ericson and V. M. Chinchilli, *Association of vitamin A deficiency with early childhood stunting in Uganda: a population-based cross-sectional study*, PLoS One, 15(5)(2020), e0233615.

[15] A. Rafferty, P. Walthery and S. King-Hele, *Analysing change over time: repeated cross sectional and longitudinal survey data*, Cathie Marsh Institute for Social ResearchPMO Human Resource Mgmt, Employment Relations & LawSocial Statistics, (2015)

[16] A. Cvetkovic-Vega, Jorge L. Maguina, Alonso Soto, Jaime Lama-Valdivia and Lucy E. Correa López, *Cross-sectional studies*, Revista de la Facultad de Medicina Humana, 21(2021), 164-170.

[17] Mia Scholten, Patrik Midlov and Anders Halling, *A Disparities in prevalence of heart failure according to age, multimorbidity level and socioeconomic status in southern Sweden: a cross-sectional study*, BMJ Open, 12(2022), e051997.