CHEST

Cross-Sectional Studies
Strengths, Weaknesses, and Recommendations

Xiaofeng Wang, PhD; and Zhenshun Cheng, MD

Cross-sectional studies are observational studies that analyze data from a population at a single point in time. They are often used to measure the prevalence of health outcomes, understand determinants of health, and describe features of a population. Unlike other types of observational studies, cross-sectional studies do not follow individuals up over time. They are usually inexpensive and easy to conduct. They are useful for establishing preliminary evidence in planning a future advanced study. This article reviews the essential characteristics, describes strengths and weaknesses, discusses methodological issues, and gives our recommendations on design and statistical analysis for cross-sectional studies in pulmonary and critical care medicine. A list of considerations for reviewers is also provided. CHEST 2020; 158(1S):S65-S71

KEY WORDS: bias; confounding; cross-sectional studies; prevalence; sampling

General Overview of Cross-Sectional Study Design

In medical research, a cross-sectional study is a type of observational study design that involves looking at data from a population at one specific point in time. In a cross-sectional study, investigators measure outcomes and exposures of the study subjects at the same time. It is described as taking a "snapshot" of a group of individuals.[1] Unlike in case-control studies (subjects selected based on the outcome status) or cohort studies (subjects selected based on the exposure status), the subjects in a cross-sectional study are simply chosen from an available population of potential relevance to the study question. There is no prospective or retrospective follow-up. Once the subjects are selected, the investigators will collect the data and assess the associations between outcomes and exposures. Figure 1 presents a schematic representation of a typical cross-sectional study.

Cross-sectional studies have been mainly used to understand the prevalence of a disease in clinical research. Prevalence refers to the proportion of persons in a population who have a particular disease or attribute at a given time, regardless of when they first developed the disease. It is important to distinguish prevalence from incidence. Incidence refers to the number of new cases that develop in a given period of time. In a cross-sectional study, researchers typically describe the distribution of variables in a population. They may assess the prevalence of a disease or association of an exposure to an outcome in a population.

In a simple hypothetical example of a cross-sectional study, we record the prevalence of COPD and investigate the association between COPD and smoking status in adult

AFFILIATIONS: From the Department of Quantitative Health Sciences (Dr Wang), Lerner Research Institute, Cleveland Clinic, Cleveland, OH; and the Department of Respiratory Medicine (Dr Cheng), Zhongnan Hospital of Wuhan University, Wuhan, China.

Both authors contributed equally to this work.

CORRESPONDENCE TO: Xiaofeng Wang, PhD, Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic, 9500 Euclid Ave/JJN3-01, Cleveland, OH 44195; e-mail: wangx6@ccf.org

DOI: https://doi.org/10.1016/j.chest.2020.03.012

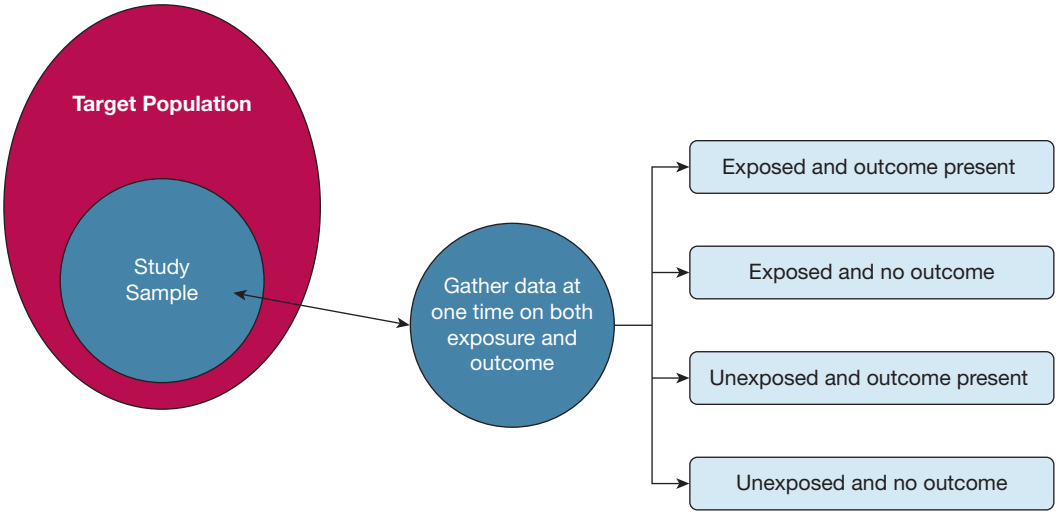

Figure 1 – *A schematic representation of a typical cross-sectional study. Data are collected on both outcomes and exposures of the individuals at a given point in time.*

patients. The outcome variable is the presence or absence of COPD, and the exposure is the smoking status. This study can be conducted by interviewing participants about their smoking history and, at the same time, assessing COPD status clinically.

Because the outcome and exposure variables are measured at the same time, it is relatively difficult to establish causal relationships from a cross-sectional study. Cross-sectional studies are usually fast and inexpensive to conduct. They are suitable for generating hypotheses and may provide information about the prevalence of outcomes and exposures that informs other study designs. In this paper, we review the essential characteristics, describe strengths and weaknesses, discuss methodological issues, and give our recommendations on design and statistical analysis for cross-sectional studies.

Description of Subtypes of Cross-Sectional Studies

Cross-sectional studies can be classified as descriptive or analytical, depending on whether the outcome variable is assessed for potential associations with exposures or risk factors. Descriptive cross-sectional studies simply characterize the prevalence of one or multiple health outcomes in a specified population. In analytical cross-sectional studies, investigators collect data for both exposures and outcomes at one specific point in time for the purpose of comparing outcome differences between exposed and unexposed subjects. The exposures and outcomes are measured simultaneously; therefore, it is difficult to determine whether the exposures preceded or followed the outcomes in an analytical cross-sectional study.

In a subtype of cross-sectional study, known as the repeated (or serial) cross-sectional study, data collection is conducted on the same target population at different time points. At each time point, investigators take a different sample (different subjects) of the target population. Thus, repeated cross-sectional studies can be used for analyzing population changes over time (also known as aggregate change over time). They cannot be used to look at individual change (as in a cohort study).

Use Cases of Cross-Sectional Studies

Example 1

Thomas et al[2] conducted a descriptive cross-sectional survey on the prevalence of dysfunctional breathing in patients treated for asthma in primary care. Of the 4,381 patients aged 17 to 65 years registered with a diagnosis of asthma from the medical records of a semirural general practice, 307 (7%) met the entry criteria and were sent the Nijmegen Questionnaire for self-completion. A total of 227 questionnaires were returned after one mailing (response rate, 74%), of which 219 were suitable for analysis. The main outcome was a score $\geq$ 23 on the Nijmegen Questionnaire. In this study, the investigator found that about one-third of women and one-fifth of men had scores suggestive of dysfunctional breathing.

Example 2

Janson et al[3] performed an analytical cross-sectional study to investigate the association between passive smoking and respiratory symptoms in the European Community

Respiratory Health Survey. The analysis included data from 7,882 adults who had never smoked, from 36 centers in 16 countries. Information was gathered through a structured interview. Spirometry and methacholine challenge were performed, and total and specific IgE were measured. Regression analysis was conducted on the variables of interest to study the association between passive smoking on respiratory symptoms and lung function. The prevalence of passive smoking in the workplace varied from 2.5% to 53.8%. The study found that passive smoking was significantly associated with nocturnal chest tightness, nocturnal breathlessness, breathlessness after activity, and increased bronchial responsiveness.

Example 3

Soriano et al[4] conducted a repeated cross-sectional survey to investigate recent trends in COPD prevalence in Spain. Estudio epidemiológico de EPOC en España (IBERPOC) and the Epidemiologic Study of COPD in Spain (EPI-SCAN) were two different epidemiologic survey studies of COPD in Spain, conducted in 1997 and in 2007, respectively. The repeated cross-sectional survey allowed the authors to compare participants from IBERPOC (n = 4,030) with those of EPI-SCAN (n = 3,802). They found that COPD prevalence in the population dropped from 9.1% in 1997 to 4.5%, a 50.4% decline. The distribution of COPD prevalence according to severity also changed from 38.3% mild, 39.7% moderate, and 22.0% severe in 1997, to 85.6% mild, 13.0% moderate, and 1.4% severe in 2007.

Benefits and Downside of Cross-Sectional Studies

The main strength of cross-sectional studies is that they are relatively quick and inexpensive to conduct. They are the best way to determine the prevalence and can study the associations of multiple exposures and outcomes. The subjects are neither deliberately exposed nor treated; thus, there are seldom ethical difficulties. Many cross-sectional studies are done through questionnaires or interviews. Using questionnaires to reach a large sample of the population of interest is relatively inexpensive but can result in low response rates. Interviews are more expensive and time-consuming than using questionnaires, potentially limiting the sample size but leading to a higher response rate. The weaknesses of cross-sectional studies include the inability to assess incidence, to study rare diseases, and to make a causal inference. Unlike studies starting from a series of patients, cross-sectional studies often need to select a sample of subjects from a large and heterogeneous study population. Thus, they are susceptible to sampling bias. We highlight the strengths and weaknesses of cross-sectional studies in Table 1.

Study Subject Considerations

Sample Size Determination

Sample size determination is an important step in the design of a cross-sectional study. Sample size calculations are different for a descriptive cross-sectional survey and an analytical cross-sectional study. When conducting a descriptive cross-sectional survey, the goal is to estimate the prevalence of a particular outcome. Investigators need to provide the assumed values of the prevalence rate, p, the desired margin of error, e (sometimes called the desired precision), and the significance level. The formula can be found in Eng.[5] Note that the sample size in a descriptive study does not depend on statistical power because this concept only applies to statistical comparisons.

TABLE 1] Strengths and Weaknesses of Cross-Sectional Studies

Strengths	Relatively quick and inexpensive to conduct
	No ethical difficulties
	Data on all variables are only collected at one time point
	Multiple outcomes and exposures can be studied
	Easy for generating hypotheses
	Many findings can be used to create an in-depth research study
Weaknesses	Unable to measure the incidence
	Difficult to make a causal inference
	Associations identified might be difficult to interpret
	Unable to investigate the temporal relation between outcomes and risk factors
	Not good for studying rare diseases
	Susceptible to biases such as nonresponse bias and recall bias

If one wants to compare two prevalence rates in an analytical cross-sectional study, the commonly used sample size formula is the same as is used when designing a cohort study.[6] For example, assume that we want to compare the prevalence rates of COPD in nonsmokers and smokers in a study. We denote that the prevalence rates are p_1 and p_2 for the two study groups, respectively. The sample size is calculated based on the following statistical hypothesis:

$H_0: p_1 = p_2$ vs. $H_1: p_1 \neq p_2$

To compute the sample size, the investigators need to provide an estimate of the prevalence, the variance of prevalence estimates, a meaningful difference between those exposed and those unexposed, the significance level, and the desired power. The formula can be found in Fleiss et al.[7] Additional discussion is presented in the sample size determination article by Wang and Ji[8] included in this supplemental issue of *CHEST*. An online calculator has been developed to help readers perform the sample size estimation, which can be found at http://riskcalc.org:3838/samplesize/.

Sampling

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. There are two major categories of sampling methods: (1) probability sampling methods, in which samples are chosen by using a method based on the theory of probability; and (2) nonprobability sampling methods, in which samples are selected based on subjective judgment. In general, probability sampling methods are preferred over nonprobability ones, as the former are considered to be more accurate and rigorous. However, in applied clinical research, there are some circumstances in which it is not feasible or practical to perform random sampling. Nonprobability sampling is applied in those situations. Martínez-Mesa et al[9] provided a useful discussion on the basic elements of selection of participants for a clinical study. Commonly used sampling methods are summarized in Table 2. Popular probability sampling techniques include simple random sampling, systematic sampling, stratified sampling, and cluster sampling. Nonprobability methods include convenience sampling, quota sampling, purposive sampling, and snowball sampling.[10] We suggest clinical investigators consult a statistician when designing a sampling strategy for a cross-sectional study.

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.[10-12] Selection bias occurs when the sample chosen

TABLE 2] Commonly Used Sampling Methods in Clinical Studies

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, *10*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

or obtained in a study is no longer representative of the overall population. It can be introduced if the selection of patients is restricted to a group with higher or lower susceptibility for developing a disease or if the exposed and unexposed groups differ in ways that predict the outcome. 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 nonresponders differ from responders. Prevalence-incidence bias (also called the Neyman bias) is also particularly common in cross-sectional studies.[13] 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.

Information bias occurs when key study variables are measured, collected, or interpreted inaccurately. 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. Table 3 displays the common types of biases and their definitions in clinical studies.

Statistical Considerations

We emphasize here a few important aspects of statistical analysis in cross-sectional studies.

Confounding

Confounding may occur in analytical cross-sectional studies when a variable is associated with the exposure and influences the outcome. 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. For restriction, investigators limit participation in the study to individuals who are similar with respect to the confounders. Stratification refers to the study of the association between exposure and outcome within different strata of the confounding variables. Propensity score matching is a statistical matching technique that entails forming matched sets of two groups of subjects who share a similar value of the propensity score.[14] Multivariable regression analysis is another way of controlling for confounding. One builds a multivariable regression model for the outcome and exposure as well as other confounding variables. Based on the regression equation, the effect of the variable of interest can be examined with confounding variables that are held constant statistically. More examples regarding confounding in clinical studies can be found in Austin[15,16] and Streiner and Norman.[17] An example of

TABLE 3] Common Types of Biases and Their Definitions in Clinical Studies

Selection bias	
Sampling bias	Some individuals within a target population are more likely to be selected for inclusion than others
Allocation bias	There is a systematic difference between participants in exposed and unexposed groups
Loss-to-follow-up bias	Some individuals lost to follow-up differ from those who were not lost to follow-up with respect to the exposure and outcome
Nonresponse bias	There is a systematic difference between responders (ie, people who complete a survey) and nonresponders (ie, people who do not complete a survey)
Prevalence-incidence bias	Also known as Neyman bias. It is a selection bias in which individuals with severe or mild disease (or both) are excluded
Information bias	
Observer bias	The investigator's prior knowledge of the disease status or treatment of the subject leads the researcher to ask questions or assess the subject differently
Interviewer bias	The tendency of the interviewer to obtain answers that support preconceived notions
Recall bias	Participants recall information on exposure differentially depending on their outcome status or recall information regarding their outcome dependent on their exposure
Detection bias	Systematic differences between groups in how outcomes are determined

confounding and a discussion of these statistical techniques can be found in the cohort study design article by Wang and Kattan[18] included in this supplemental issue of *CHEST*. Further discussion of confounding can be found in the directed acyclic graphs article by Etminan et al[19] also included in this supplemental issue of *CHEST*.

Modeling

In analytical cross-sectional studies, investigators may develop explanatory regression models or diagnostic prediction models. In an explanatory model, variables that have a scientifically meaningful and statistically significant relationship with an outcome are identified. In a diagnostic model, multiple predictors are combined to estimate the probability that a particular condition or disease is present at the moment of prediction. Note that diagnostic models are different from prognostic models in cohort studies, which are usually longitudinal.[20] Variable selection is vital in the process of model building. Stepwise selection methods and *P* value-based criteria are discouraged due to overfitting and because they often allow too many parameters to be included, which reduces the generalizability of the model. Moons et al[20] suggested that backward elimination with the Akaike information criterion is preferred if there is no way to avoid automated variable selection. Modern shrinkage or penalization procedures, such as LASSO/least absolute shrinkage and selection operator, elastic net, and their variants, are recommended for the study with rare events or with a large number of predictors. Diagnostic prediction models should include some form of internal validation, such as cross-validation or bootstrapping, particularly in the situation that no additional external validation is performed. Additional details are given in the prediction modeling article by Kattan and Gerds[21] included in this supplemental issue of *CHEST*.

Reporting Considerations

Only with full and transparent reporting of information on all aspects of a cross-sectional study can potential usefulness of its findings and risk of bias be adequately assessed. We suggest that investigators report their cross-sectional studies following the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement, which contains a checklist of 22 items that are considered essential to report.[22] In the STROBE initiative's explanation and elaboration paper,[23] several examples are used to guide how to improve the reporting of observational studies.

When diagnostic prediction models are developed in a cross-sectional study, we recommend that investigators follow the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) statement.[24] The statement is a checklist of 22 items deemed essential for transparent reporting a prediction model study. It is an explanation and elaboration document[20] that describes the rationale, clarifies the meaning of each item, and provides a valuable reference of issues to consider when reporting the design, conduct, and analysis of prediction model studies.

Short List of Questions to Guide the Reviewer

When reviewing a cross-sectional study, a reviewer should consider commenting on the following:

1. **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?
2. **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?
3. **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?

Acknowledgments

Financial/nonfinancial disclosures: None declared.

References

1. Carlson MDA, Morrison RS. Study design, precision, and validity in observational studies. *J Palliative Med*. 2009;12(1):77-82.
2. Thomas M, McKinley RK, Freeman E, Foy C. Prevalence of dysfunctional breathing in patients treated for asthma in primary care: cross sectional survey. *BMJ*. 2001;322(7294):1098-1100.
3. Janson C, Chinn S, Jarvis D, et al. Effect of passive smoking on respiratory symptoms, bronchial responsiveness, lung function, and

total serum IgE in the European Community Respiratory Health Survey: a cross-sectional study. *Lancet*. 2001;358(9299):2103-2109.

4. Soriano JB, Ancochea J, Miravitlles M, et al. Recent trends in COPD prevalence in Spain: a repeated cross-sectional survey 1997-2007. *Eur Respir J*. 2010;36(4):758-765.
5. Eng J. Sample size estimation: how many individuals should be studied? *Radiology*. 2003;227(2):309-313.
6. Fleiss JL, Levin B, Paik MC. *Statistical Methods for Rates and Proportions*. 3rd ed. New York, NY: John Wiley & Sons; 2013.
7. Fleiss JL, Tytun A, Ury HK. A simple approximation for calculating sample sizes for comparing independent proportions. *Biometrics*. 1980;36(2):343-346.
8. Wang X, Ji X. Sample size estimation in clinical research: from randomized controlled trials to observational studies. *Chest*. 2020;158(suppl 1):S12-S20.
9. Martínez-Mesa J, González-Chica DA, Duquia RP, Bonamigo RR, Bastos JL. Sampling: how to select participants in my research study? *Anais brasileiros de dermatologia*. 2016;91(3):326-330.
10. Martin SW, Meek AH, Willeberg P. *Veterinary Epidemiology: Principles and Methods*. Ames, IA: Iowa State University Press; 1987.
11. Merrill RM. *Introduction to Epidemiology*. 7th ed. Burlington, MA: Jones & Bartlett Publishers; 2015.
12. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. 3rd ed. Philadelphia, PA: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2012.
13. Tripepi G, Jager KJ, Dekker FW, Zoccali C. Selection bias and information bias in clinical research. *Nephron Clin Pract*. 2010;115(2):c94-c99.
14. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70(1):41-55.
15. Austin PC. A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Stat Med*. 2008;27(12):2037-2049.
16. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res*. 2011;46(3):399-424.
17. Streiner DL, Norman GR. The pros and cons of propensity scores. *Chest*. 2012;142(6):1380-1382.
18. Wang X, Kattan MW. Cohort studies: design, analysis, and reporting. *Chest*. 2020;158(suppl 1):S72-S78.
19. Etminan M, Collins GS, Mansournia MA. Using causal diagrams to improve the design and interpretation of medical research. *Chest*. 2020;158(suppl 1):S21-S28.
20. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med*. 2015;162(1):W1-W73.
21. Kattan MW, Gerds TA. A framework for the evaluation of statistical prediction models. *Chest*. 2020;158(suppl 1):S29-S38.
22. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Int J Surg*. 2014;12(12):1495-1499.
23. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med*. 2007;4(10):e297.
24. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med*. 2015;162(1):55-63.