


Research Design: Cohort Studies

Chittaranjan Andrade¹ 

ABSTRACT

In a cohort study, a group of subjects (the cohort) is followed for a period of time; assessments are conducted at baseline, during follow-up, and at the end of follow-up. Cohort studies are, therefore, empirical, longitudinal studies based on data obtained from a sample; they are also observational and (usually) naturalistic. Analyses can be conducted for the cohort as a whole or for subgroups amongst which comparisons can be drawn. Because there is no randomization to the subgroups of interest, cause and effect relationships cannot be determined, and relationships between variables must be stated as associations that may or may not be influenced by confounding. The cohort that is studied can be prospectively or retrospectively defined, and each method has its advantages and disadvantages. These and other issues are explained with the help of examples. A special note is made of cohort studies in Indian psychiatry.

Keywords: Cohort study, research design, prospective study, retrospective study, STROBE guidelines, India

Previous articles in this series on research design discussed classifications in research design¹ and prospective and retrospective, cross-sectional and longitudinal studies.² This article examines a specific research design that is increasingly being employed in medicine and psychiatry: the cohort study.

Cohort Study: General Description

A cohort is a group of subjects. In a cohort study, the cohort is made up of subjects who meet the study selection criteria. Identification of the cohort, or recruitment, occurs across a period of time. The cohort so identified is followed for a

further period of time. The study usually ends on a set date or when the desired endpoint has been reached. Assessments are conducted at recruitment, during follow-up, and at the end of follow-up.

Examples of Cohort Studies

1. All children born in a hospital during a two-year period are followed until each child reaches the age of 18 years; the main objective of the study is to examine the incidence and predictors of different childhood-onset psychiatric disorders.
2. All patients newly diagnosed with schizophrenia are followed for 10 years to examine how different socio-demographic (e.g., male vs. female),

clinical (e.g., short vs. long duration of untreated psychosis), and treatment (e.g., first vs. second generation antipsychotics) characteristics influence the course and outcome of the disorder.

3. All doctors in a geographical area are followed until death to examine how their lifestyle behaviors and professional exposures influence their risk of developing cancer, cardiovascular disease, and dementia.

The Framingham Heart Study, the Nurses Health Study, and the Nun Study are examples of well-known cohort studies. In the field of mental health, the Adolescent Brain Cognitive Development (ABCD) study is a large ongoing cohort study that will follow a cohort of >10,000 children from childhood to adult life; the purpose of the study is to understand the determinants of cognitive, social, emotional, and physical development.³

Characteristics of Cohort Studies

Examined from the perspective of research design, cohort studies are empirical because they collect and examine data. They are sample-based because a group of individuals is studied. They are always longitudinal because there is a follow-up, but can be prospectively

¹Dept. of Clinical Psychopharmacology and Neurotoxicology, National Institute of Mental Health and Neurosciences, Bangalore, Karnataka, India.

HOW TO CITE THIS ARTICLE: Andrade C. Research Design: Cohort Studies. *Indian J Psychol Med.* 2022;44(2):189–191.

Address for correspondence: Chittaranjan Andrade, Dept. of Clinical Psychopharmacology and Neurotoxicology, National Institute of Mental Health and Neurosciences, Bangalore, Karnataka 560029, India. E-mail: andradec@gmail.com

Submitted: 27 Dec. 2021
Accepted: 29 Dec. 2021
Published Online: 22 Feb. 2022



Copyright © The Author(s) 2022

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution- NonCommercial 4.0 License (<http://www.creativecommons.org/licenses/by-nc/4.0/>) which permits non-Commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

ACCESS THIS ARTICLE ONLINE
Website: journals.sagepub.com/home/szj
DOI: 10.1177/02537176211073764

or retrospectively defined, as was clarified in an earlier article.³ There is no randomization to subgroups of interest and no blinding of subjects or raters to the characteristics of interest. They can be uncontrolled, with outcomes examined in a group of persons defined by a single characteristic (see example 1 in the previous section), or they can be quasi-controlled, with subgroups being compared (see example 2 in the previous section).⁴ They are usually observational and naturalistic, though interventions can be factored into the study protocol. For readers unfamiliar with these terms, “observational” means that the investigators merely record data; they do not intervene in patient care. “Naturalistic” means that patients receive whatever treatment their health care professionals consider necessary and appropriate, or, in other words, treatment as usual.

Retrospective Cohort Studies

As an example of a retrospective cohort study, a cohort can be defined to comprise all children born in a health care system between 1980 and 1990. The health care system can be a single hospital, a group of hospitals, an insurance database, or any other database that maintains records of medical information. The data of the children identified are examined and extracted from paper or electronic charts until, say, 2020, to ascertain what maternal, gestational, and early postnatal characteristics, recorded during and after the pregnancy, predict adult mental health.

If large electronic health care databases exist, as they do, e.g., in the Scandinavian countries and the USA, a very large cohort, with tens or hundreds of thousands of subjects, can quickly be identified and followed up in the database. Information on the variables of interest can be extracted in a relatively short period of time with little effort and expense. Large cohorts, so identified, will have high statistical power to examine hypotheses of interest.

Retrospectively defined cohorts are, however, compromised in quality because chart data or health care database data tend to be casually rather than accurately recorded; different health care personnel would have obtained

and recorded the data for different subjects in the cohort with potentially poor inter-rater reliability, and data for important independent variables may be unavailable for many or all subjects. These advantages and disadvantages are common to all retrospective studies.

Prospective Cohort Studies

As an example of a prospective cohort study, pregnant women can be recruited across the course of two years; relevant participant and gestational data can be recorded, and the children who are born can be followed for, say, the next 30 years to determine what maternal, gestational, and early postnatal characteristics, recorded during and after the pregnancy, predict adult cognitive and mental health.

An advantage of prospective cohort studies is that all relevant variables can be thought of in advance, and data related to these variables can be accurately measured and recorded by trained study staff. Disadvantages are that prospective cohort studies are expensive to conduct and take long to complete; in fact, the investigators who analyze the data may be the successors of those who started the study. For practical reasons related to expense and effort, prospective cohorts are mostly smaller than retrospective cohorts. These advantages and disadvantages are common to all prospective studies.

Theoretical and Practical Issues

Cohort studies are complicated by many issues. Characteristics that are recorded at the baseline, such as medications and drug doses, or smoking and drinking variables, or dietary variables, may change during the course of the study. Repeated follow-up visits need to be scheduled. Subjects may drop out for various reasons, including death. Different subjects are followed up for different durations of time. Most importantly, when comparing subgroups of interest, such as outcomes in children gestationally exposed vs. unexposed to antidepressant drugs, because subjects were not randomized to their respective groups, many confounding variables can complicate the analysis of data and the interpretation of the analysis. To

overcome these problems, subjects can be propensity score-matched to reduce confounding,⁵ and Cox proportional hazards regression, which yields a hazard ratio, can be used to adjust for confounders and to take into consideration the varying duration of follow-up as well as the time of occurrence of the outcome of interest. Whereas randomized controlled trials (RCTs) follow the CONSORT guidelines and systematic reviews and meta-analyses follow the PRISMA guidelines, cohort studies follow the STROBE guidelines.

Cohort studies can generate useful epidemiological data. However, when examining relationships between variables, because there is no randomization, the results of analyses can only be considered as associations. As an example, gestational exposure to antidepressant drugs may be associated with the development of autism spectrum disorder in the offspring, but whether the association represents a cause-effect relationship cannot be stated. Cause and effect relationships, as studied in RCTs, cannot be determined from observational data generated in cohort studies. However, cohort studies can be used to examine hypotheses that cannot be examined in RCTs, as in the example cited above; this is because RCTs are hard to conduct in conditions such as pregnancy. Cohort studies also generate large quantities of data that can be studied from different perspectives.

Implications for Research in India

Long-term funding is hard to obtain in India, and so few cohort studies in psychiatry have been published from this country. An outstanding example of a prospective cohort study is the Madras Longitudinal Study, initiated in 1981.⁶ Another important example is the more recent Thirhalli Cohort.⁷ There are many other examples of prospective cohorts, such as the Prospective Assessment of Maternal Mental Health Study,⁸ but these are considerably shorter studies. India does not yet have a digitized health care structure that would allow retrospective cohorts to be constructed.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Chittaranjan Andrade  <https://orcid.org/0000-0003-1526-567X>

References

1. Andrade C. Describing research design. *Indian J Psychol Med* 2019; 41(2): 201–202.
2. Andrade C. Simultaneous descriptors of research design. *Indian J Psychol Med* 2021; 43(6): (in press).
3. Jernigan TL, Brown SA; ABCD Consortium Coordinators. Introduction. *Dev Cogn Neurosci* 2018; 32: 1–3.
4. Andrade C. The limitations of quasi-experimental studies, and methods for data analysis when a quasi-experimental research design is unavoidable. *Indian J Psychol Med* 2021; 43(5): 451–452.
5. Andrade C. Propensity score matching in nonrandomized studies: A concept simply explained using antidepressant treatment during pregnancy as an example. *J Clin Psychiatry* 2017; 78(2): e162–e165.
6. Rangaswamy T and Cohen A. Invited commentary from a LAMIC country: Thirty-five years of schizophrenia—the Madras Longitudinal study. *Schizophr Res* 2020; 220: 27–28.
7. Bagewadi VI, Kumar CN, Thirthalli J, et al. Standardized mortality ratio in patients with schizophrenia—findings from Thirthahalli: A rural south Indian community. *Indian J Psychol Med* 2016; 38(3): 202–220.
8. Chandra PS, Bajaj A, Desai G, et al. Anxiety and depressive symptoms in pregnancy predict low birth weight differentially in male and female infants—findings from an urban pregnancy cohort in India. *Soc Psychiatry Psychiatr Epidemiol* December 2021; 56(12): 2263–2274.