Column: Methods In Psychiatric Research

COHORT STUDIES

Pankajakshan Vijayanthi Indu[1], K Vidhukumar [2]*

[1]Government Medical College, Kozhikode
[2]Government Medical College, Thiruvananthapuram
* Corresponding Address: Additional Professor in Psychiatry, Government Medical College, Thiruvananthapuram-695011, Kerala. Email: kumarv68@gmail.com

ABSTRACT

Cohort studies are observational analytical studies and are placed just below randomized controlled trials in the hierarchy of scientific evidence. They are also called longitudinal studies and can be prospective, retrospective or ambidirectional. In cohort studies, groups (cohorts) that are identical in every aspect other than the exposure are followed up for the outcome. Critical components of a cohort study are the definition of the objective/s, choice of the study population and comparison group, measurement of exposures and outcomes, follow-up, analysis and interpretation. Assembling cohorts, accurate measurement of exposure and outcome, and addressing dropout are the main design imperatives in cohort studies. The advantages and disadvantages of cohort studies and an overview of analysis are also presented.

Keywords: cohort study, nested case-control study, case-cohort study, incidence

Cohort studies are observational analytical studies in which a group or groups of individuals are followed up in time to detect how many develop an outcome of interest like disease, death, an event or a change in health status or behaviour. A cohort refers to a group of persons who share a common characteristic. For e.g., doctors, nurses, smokers, people born in the same year, etc.[1] Etymologically, the term "cohort" stems from the Latin word "*cohors*"—which refers to a military unit, one of the 10 divisions of a Roman legion. Thus, a cohort constitutes a group of persons marching forward in time from exposure to one or more outcomes. Various terms are used synonymously to refer to cohort studies: incidence, longitudinal, prospective, follow-up, concurrent, or forward-looking studies.[2] Cohort studies are useful study designs to study incidence, causes and prognosis.[3]

Types of cohort studies

There can be either prospective or retrospective cohort studies. A third type is ambidirectional cohort studies.[2] In prospective cohort studies, a group of people in whom exposure is assessed at baseline is followed up in time to see whether they develop an outcome of interest. In a retrospective cohort study (also known as historical cohort study or non-concurrent prospective study), when the researcher starts the study, the follow-up has already been completed—both the exposure and outcome of interest have already occurred, and the corresponding data has been collected for some other purpose. The researcher identifies a cohort of eligible subjects, and the data regarding exposure and outcome during the period of observation are used for analysis.[3,4] In ambidirectional cohort studies, data collection occurs retrospectively as well as prospectively. Both short- and long-term outcomes can be studied by this design.[2] For instance, to study the incidence of multiple births as an outcome of assisted reproductive technologies, a prospective or retrospective design can be used. To study the association of assisted reproductive technologies with multiple births and ovarian cancer later in life, the researcher can look back at past medical

Access the article online:
https://kjponline.com/index.php/kjp/article/view/319
DOI: https://doi.org/10.30834/KJP.34.2.2022.319
Received: 3/01/2022. Web publication: 09/01/2022

QR Code

Please cite this article as: PV Indu, K Vidhukumar. Cohort studies (Column: Research Methods in Psychiatry). Kerala Journal of Psychiatry 2021;34(2): 164-169

records of such a group of women to study multiple births and follow-up the same group for later development of ovarian cancer in an ambidirectional design.[2] (See Figure 1.)

Two study designs (cohort and case-control) are combined in a nested case-control study. Cases are identified retrospectively or prospectively from a cohort, and controls are selected—usually matched to the cases—from the remainder of the cohort when the cases develop the disease. A case-cohort design is a variant of the nested case-control study, in which a sub-cohort is randomly drawn from the full cohort at the start of the study. The case-cohort sample includes this sub-cohort and all those who develop the outcome of interest (the cases) from the entire cohort.[5,6]

Cohort studies are similar to intervention studies in that groups are selected based on the exposure status and followed up for the outcome of interest. However, unlike the latter, the researcher is not allocating the exposure in the former.[7] (See Figure 2.)

The issues to focus on while designing a cohort study are:

- Definition of the objective/s
- Choice of the study population and the comparison group
- Measurement of exposure
- Measurement of outcomes
- Follow-up
- Analysis and
- Interpretation[7]

Definition of the objective/s

As in any other research, the research hypothesis has to be formulated clearly at the outset. The exposure/s and outcome/s have to be defined explicitly. Generally, these studies are designed based on hypotheses generated from the evidence obtained from cross-sectional or case-control studies.[7]

Choice of study population and comparison group

Study population

In cohort studies, one group of participants are usually assessed at baseline for one or more exposures, divided into exposure categories and then followed up in time—e.g., cases of first-episode schizophrenia. Or else, two separate groups can be selected based on having or not having the exposure and followed up—e.g., smokers vs non-smokers.[4] The cohort can be drawn from a general population group like the Framingham Heart Study, where participants were selected from a geographically well-defined area. Multiple exposure factors can be studied in this case. Alternatively, they may be chosen from a narrowly defined population, like specific socio-professional groups (e.g., nuns in the Nun study of ageing and Alzheimer's disease). When the exposure is rare, a highly exposed group would be preferable to the general population. For instance, the Life Span Study assessed the long-term health hazards of ionizing radiation in the survivors of atom bomb explosions in Japan.[7]

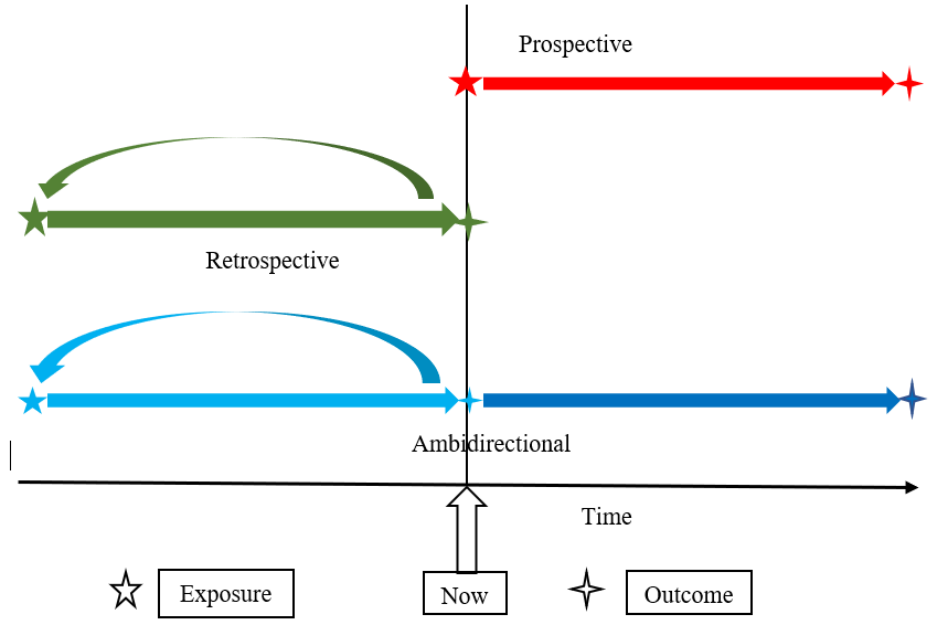
Comparison group

Selecting an appropriate comparison group of unexposed persons is crucial in designing a cohort study. The unexposed group must be as similar as possible to the exposed group on almost all factors related to the outcome of interest, except the exposure.[1,7] If so, the comparison group will reflect the background rate of the outcome of interest in the community.[2] The source for the comparison group can be internal or external. It is desirable to have an internal comparison group. When the cohort is from the general population or a well-defined population, members who are either unexposed or have low levels of exposure can be the comparison group. For example, in the cohort of British physicians, non-smokers were the comparison group against smokers.[7] In situations where a satisfactory internal comparison group is unavailable, an external comparison group can be selected. In cohort studies assessing the outcome of occupational exposure to radiation, those employed in similar settings without exposure to radiation can be the external comparison group. Another external comparison group can be selected from the general population also.

Measurement of exposure

Information regarding the exposure of interest can be obtained from several sources—records collected for other purposes like medical or employment records, details provided by the participants or reliable informants through interviews or questionnaires, data

Figure 1. Schematic representation of prospective, retrospective and ambidirectional cohort studies (diagram adapted from Ref. 2)



obtained from medical examination or laboratory investigation of the participants, or even measurement of the environment in which the participant lived or worked. The details of the exposure have to be obtained at baseline. Further, age at first exposure, duration or pattern of exposure and changes in exposure over time can be assessed as appropriate.[7]

In a retrospective cohort study, data regarding exposure is collected from pre-existing records. As this information had been collected for purposes other than the study, the data tends to be less accurate and detailed than if it was collected prospectively. Additional information required for the cohort study may also be lacking in these studies.[7]

Measurement of outcome/outcomes

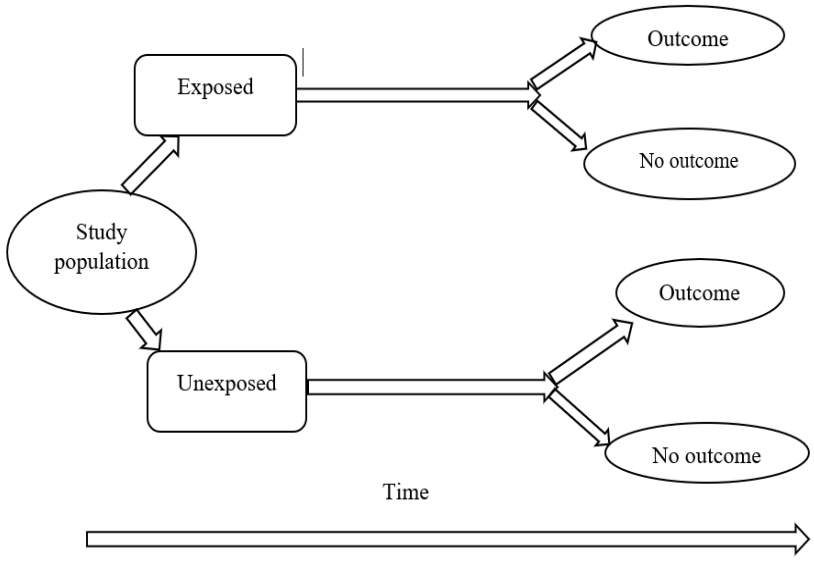
Outcomes have to be defined at the beginning of the study itself and should be clear, specific and measurable. Multiple outcomes can be studied in cohort studies. As cohort studies assess the development of the outcome of interest over time, it should not be present in the participants at the time of entry into the study. Hence, all potential participants have to be evaluated for the outcome of interest at baseline, and, if present, excluded from the study.[1,7]

The outcome/s of interest can be ascertained using existing surveillance systems like death registers or registries for particular diseases or even medical or insurance records. When no such system exists to document the outcome of interest, a system for ascertaining the outcome of interest has to be set up in the cohort itself. This can be done using self-administered or clinician-administered questionnaires, interviews, regular physical or neuropsychological evaluation of the participants, or investigations. Whatever be the method chosen to confirm the outcome of interest, it has to be used identically for the exposed and the comparison groups. To avoid measurement bias, it would be ideal to blind the interviewer or person ascertaining the outcome of interest to the exposure status of the participant.[2,7]

Follow-up

Methods to ensure follow-up of the participants have to be described clearly before beginning the study. Periodic contacts with the participants can be ensured through home visits, mailed questionnaires, telephone calls or online methods. Following up a large number of participants for prolonged periods is a costly endeavour. Existing surveillance systems can be used to trace the

Figure 2. Prospective cohort study



participants at a lesser cost. The exit criteria should also be specified at the outset; it could be the end of the follow-up period or the occurrence of the outcome of interest.[7]

Analysis

As cohort studies are used to estimate incidence, risk and prognosis, the first step in the analysis is assessing the incidence of the outcome of interest.

Incidence

Incidence refers to the number of new cases that develop in a population at risk during a specified period of time. There are two measures of incidence—Cumulative Incidence (CI) and Incidence rate or Incidence Density (ID).[8]

CI assumes that the entire population at risk at the beginning is followed up for the specified period of time for the occurrence of the outcome of interest. It refers to the proportion of people who develop a disease or outcome of interest during a specified period of time and is calculated as follows:

$$\text{CI} = \frac{\text{Number of new cases of a disease or outcome during a given period of time}}{\text{Total population at risk}}$$

As participants are followed up for different periods of time, the calculation of incidence can be restricted to the period during which the entire population provided information. This can be achieved by calculating the sum of the time during which each person remained under observation and remained disease-free, i.e., person-time, which could be person-days, person-months or person-years. ID refers to the incidence during a period of time when the entire study population provided information and is calculated as follows:

$$\text{ID} = \frac{Number\ of\ new\ cases\ during\ a\ period\ of\ time}{Total\ person{-}time\ of\ observation}$$

Measure of association

The incidence of the outcome in the exposed and unexposed groups can be compared to obtain a measure of the association between the exposure and the outcome. Relative risk or risk ratio (RR) gives the magnitude of association between the exposure and outcome in a cohort study. It is the ratio of the incidence of the outcome in the exposed group (I_e) to the incidence of the outcome in the unexposed group (I_0). See Table 1. This applies when CI is estimated. If ID is estimated in a study, the rate ratio can be calculated as the ratio of incidence rate among exposed to the incidence rate among unexposed.

difference or attributable risk, attributable risk percentage, population attributable risk and population attributable risk percentage can be calculated from the data obtained from cohort studies.[8]

Table 1. Calculation of Relative Risk from a cohort study

	Cases	Controls	
Exposure	a	b	a + b
No exposure	c	d	c + d
	a + c	b + d	

$$I_e = CI_e = a/(a+b)$$

$$I_0 = CI_0 = c/(c+d)$$

$$RR = I_e / I_0 = a/(a+b) \div c/(c+d)$$

Advanced analysis

Survival or time-to-event may be the outcome studied in cohort studies. So also, repeated measures of biomarkers may be evaluated to assess disease progression or prognosis. Multivariate analytical methods like survival analysis, time-to-event analysis or longitudinal data analysis can be used to bring out the factors related to survival or trajectories of markers of disease progression.[5] Model building is endeavoured in cohort studies, both explanatory and predictive. Explanatory modelling identifies variables with a scientifically meaningful and statistically significant relationship with an outcome. In contrast, predictive modelling attempts to predict the probability of an outcome (diagnosis or prognosis) for an individual. Advanced statistical techniques are used for these analyses.[6]

Interpretation

The role of chance, bias and confounding have to be considered while interpreting the results of a cohort study. The interpretation of RR is much the same as that for OR, which was discussed in the previous edition. The role of chance can be understood from the p-value and 95% confidence interval.[9]

Bias

Being aware of the exposure status of the participants can introduce measurement bias in the assessment of the outcome. Blinding the evaluator to the exposure status can help prevent this.[7] Loss to follow-up due to dropouts or death is a source of potential bias in cohort studies. As a rule of thumb, loss to follow-up should not exceed 20% of the study sample. Minimizing loss to follow-up is the best method to prevent this bias.[6]

Confounding

This is a major factor that can affect the interpretation of the results of a cohort study. All potential confounding variables have to be considered in the design stage of the study and either excluded or measured. Only if data regarding these variables is available it can be adjusted statistically during the analysis. In retrospective cohort studies, data regarding potential confounders is usually not available, affecting the interpretation of the findings.[7]

Advantages of cohort studies

- Especially useful to study rare exposures
- Multiple outcomes of a single exposure can be studied
- Temporal relationship between exposure and outcome can be established
- In prospective cohort studies, bias in the measurement of exposure and recall bias can be minimized
- Incidence of outcome in exposed and unexposed groups can be assessed[7,8]

Disadvantages of cohort studies

- Not efficient to evaluate rare outcomes
- Prospective studies are extremely expensive and time-consuming
- Retrospective studies may have a dearth of adequate records and data
- Losses to follow-up can affect the validity of the findings
- Changes in exposure status and diagnostic criteria over time can affect the exposure/outcome status[7,8]

Conclusion

While designing a cohort study, the research question, the study cohort and exposure and outcome variables

have to be clearly defined. The potential for selection, measurement or misclassification biases must be considered and addressed in the design phase itself. Exposure status and its change over time and outcome have to be evaluated meticulously. Potential confounders must be identified and controlled for in the design and/or analysis. Follow-up must be ensured, minimizing loss to follow-up. Missing data should be addressed during the analysis of the results. While reporting the findings of cohort studies, the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines can be used as a checklist.[6]

References

1. Alexander LK, Lopes B, Ricchetti-Masterson K, Yeatts KB. Cohort Studies. ERIC Notebook [Internet]. Second. 2015 [cited 2021 Dec 28]. p. 6. Available from: https://sph.unc.edu/files/2015/07/nciph_ERIC5.pdf.
2. Grimes DA, Schulz KF. Cohort studies: marching towards outcomes. Lancet (London, England). 2002 Jan 26;359(9303):341–5.
3. Mann CJ. Observational research methods. Research design II: cohort, cross sectional, and case-control studies. Emerg Med J. 2003 Jan 1;20(1):54–60.
4. Euser AM, Zoccali C, Jager KJ, Dekker FW. Cohort Studies: Prospective versus Retrospective. Nephron Clin Pract. 2009 Oct;113(3):c214–7.
5. Samet JM, Muñoz A. Evolution of the cohort study. Epidemiol Rev. 1998;20(1):1–14.
6. Wang X, Kattan MW. Cohort Studies: Design, Analysis, and Reporting. Chest. 2020 Jul 1;158(1S): S72–8.
7. Silva I dos S. Cohort studies. In: Cancer Epidemiology: Principles and Methods. Lyon, France; International Agency for Research on Cancer; 1999. p. 165–88.
8. Hennekens CH, Buring JE MS. Epidemiology in Medicine. First Edit. Boston: Little Brown; 1987.
9. Indu PV, Vidhukumar KP. Case-Control Studies. Kerala J Psychiatry 2021 Aug 3;34(1):84–7.