# How to Conduct and Write a Cohort Study

Michael Ian N. Sta. Maria, MD, FPAFP and Nicolas R. Gordo Jr., MD, MHA

Cohort studies is an epidemiologic study that follows a group of individuals who share a common characteristic at the start of the study to observe the emergence of outcomes. Cohort studies are classified based on the population characteristics from where they were drawn, the way the data collection occurred or if its open or closed. This allows the computation of the absolute risk or the incidence of an outcome.

There are several advantages in conducting a cohort study, such as clarity of temporal relationship of the exposure and outcome, permits the computation of incidence, permits multiple effects of a single exposure, and avoids selection bias on admission. While there are advantages, there are also disadvantages in doing this study, such as it requires long follow-up, need of large sample size, maybe costly, and may make it difficult to argue causation due to the presence of confounding.

The statistical test that can be used to analyze the results will depend on the type of variable used. Statistical test such as T-test, Chi square test, and Regression can be used.

Writing the final report follows the STROBE guidelines.

**Definition**

Cohort studies are epidemiological investigations that track the emergence of outcomes among a group of individuals (the cohort) who share a common characteristic at the start of the study. Furthermore, the terms incidence study and follow-up may be used.[1-4] Cohort studies can be further classified according to the population characteristics from which they were drawn (e.g., birth cohort), the way data collection occurred (i.e., prospective, retrospective, or ambispective), or whether the cohort is open or close.[5]

Measurements of exposure variables, including time-varying characteristics and covariates, are done. The cohort is followed up on, and the incidence of outcome variable(s) is observed. The study may employ specified and measured variables of interest to compare the incidence of new disease cases (or other outcomes) between groups that have been exposed and those that have not (refer to Figure 1).[2] Cohort studies, like case-control studies, are longitudinal in nature since exposure and disease data span multiple time periods. The difference is prior to commencing a case-control study, an assessment is conducted on the outcome variable then participants are followed up to observe for exposure variables. Compared to randomized controlled trials, cohort studies do not employ random assignment of exposure. Conversely, exposure status can be obtained through voluntary actions (e.g., smoking) or by chance (e.g., genetic polymorphisms).

Based on the exposure status the outcome (s) is measured, which is typically the incidence of disease (or other outcome) over a specified time (absolute risk). As a result, contingent upon its intended application, the comparison of absolute risk between exposure statuses can be delineated as attributable risk (risk difference), population-attributable risk, relative risk (risk ratio), or population-attributable fraction (refer to Table 1).[2]

Cohort studies are considered the most advanced form of observational study due to their ability to provide empirical evidence regarding exposure's impact on prevention, risk, prognosis, and treatment. The direction of relationships between variables can be precisely characterized in cohort studies due to the temporal order of variable assessments; causes must occur before effects.

**Advantages of Cohort Studies**

*Clarity of Temporal Relationship*

The cohort offers substantial knowledge into causality and the direct assessment of risk associated with an outcome. Given that exposure was assessed prior to the manifestation of the outcome, there is an evident temporal relationship between exposure and outcome. Determining the temporal relationship between the exposure and the outcome is an essential determinant of causation. Case-control studies involve to-be-determined individuals who have experienced the outcome and those who have not, followed by an assessment of their prior exposures. Compared to case-control studies, one begins with individuals with the outcome and those without and then determines their previous exposures.

**Table 1**. Measures of effect.

| Measure | Definition | Formula |
|---|---|---|
| Absolute Risk | The probability of an event in a population under study. The value is the same as incidence. | $I = \frac{\text{Number of new cases in a specified period}}{\text{Number of people in the group}}$ |
| Relative Risk (Risk ratio) | The ratio of the absolute risk in the exposed to unexposed group. | $RR = \frac{I_{Exposed}}{I_{Nonexposed}}$ |
| Attributable Risk (Risk difference) | The amount of risk of the outcome in the exposed group that can be attributed to the exposure. | $AR = I_{Exposed} - I_{Nonexposed}$ |
| Population-attributable risk | The product of the attributable risk and the prevalence of exposure to the risk factor in a population. Measures the excess disease incidence in a community associated with a risk factor. | $AR_P = AR \times P$ |
| Population-attributable fraction | The fraction of disease occurrence in a population is associated with a particular risk factor. | $AF_P = \frac{AR_P}{I_T}$ |

*Where I = incidence; P = prevalence of exposure to a risk factor; AR = attributable risk; $AR_P$ = population-attributable risk; $AF_P$ = population-attributable fraction; $I_T$ = total incidence of disease in a population.
Source: Bonita et al., 2006

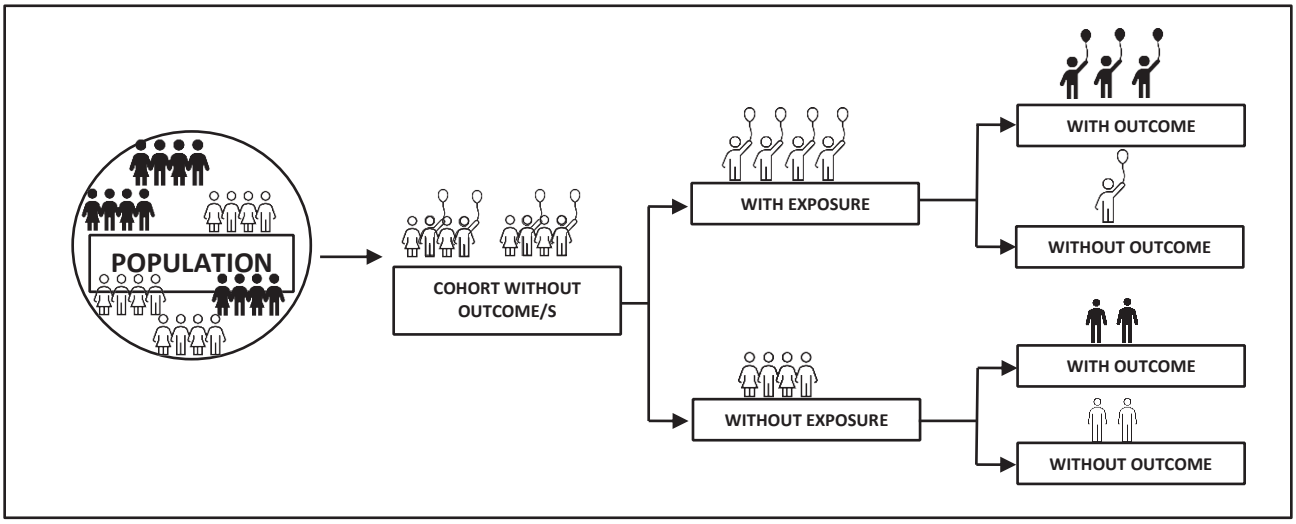


**Figure 1**. Illustration on how a cohort study is conducted.

*Permit Computation of Incidence*

Cohort studies facilitate the estimation of disease (outcome) incidence within exposure groups (absolute risk).

*Promote Research into Rare Exposures*

While a cohort design can be used to investigate common exposures (e.g., cardiovascular disease and cancer risk factors in the Nurses' Health Study), it is beneficial for evaluating the effects of rare or unusual exposures because researchers can identify enough subjects who have been exposed to the rare or unusual exposure.

*Permit Examination of Multiple Effects of a Single Exposure*

Multiple outcomes that may be associated with the exposure of interest may be observed in cohort studies.

*Avoid Selection Bias in Admissions*

Cohort studies, especially prospective cohort studies, reduce the likelihood that the results will be biased by selecting subjects for the comparison group who may be more or less likely to have the outcome of interest because the outcome is unknown at baseline when exposure status is determined. Nevertheless, selection bias can

occur in retrospective cohort studies (because outcomes have already happened at the time of selection) and prospective cohort studies due to differential loss of follow-up.

**Disadvantages of Cohort Studies**

*Requires Long Period of Follow-up.*

To account for the potential delay in the manifestation of most outcomes following exposure, extended periods of follow-up may be required. As illustrated in Figure 2, the induction and latency period for human papillomavirus-induced cervical cancer span several years. Consequently, it might be imperative to conduct prolonged follow-ups with study participants. This is a possible occurrence in prospective cohort studies.

The induction period refers to the temporal span that transpires from the moment "sufficient cause" is established until the initial pathological alterations of the disease manifest within the organism. The latency period refers to the duration that passes from the onset of pathological transformation until its detection of disease.

*May Need a Large Sample Size*

This is the case when the observed outcome is uncommon in both the exposed and unexposed groups; therefore, cohort studies are not recommended for rare outcomes; case-control studies are preferable.

*Maybe Costly*

As a result of the prolonged duration and substantial sample size.

*May Make it Difficult to Argue Causation*

Due to the presence of confounding variables and the possibility that other events that transpired in the intervening time period influenced the outcome; causality may be difficult to establish in cohort studies.

**Types of Cohort Study According to Timing of Data Collection**

*Prospective Cohort Studies*

Prospective cohort studies are those in which the investigation commences with the determination of the exposure status of subjects, followed by the subsequent assessment of the outcome (refer to Figure 3).[3-7]

Advantages

1. Considerations pertaining to significant and pertinent confounding variables that vary over time.
2. Prospects to investigate the causes of diverse outcomes.

Disadvantages

1. Potentially required to monitor a multitude of subjects for an extended period of time.
2. They can be time-consuming and expensive.
3. They are unsuitable for uncommon occurrences.
4. They are inappropriate for conditions characterized by protracted induction and latency periods.
5. Differential loss to follow-up can introduce bias.

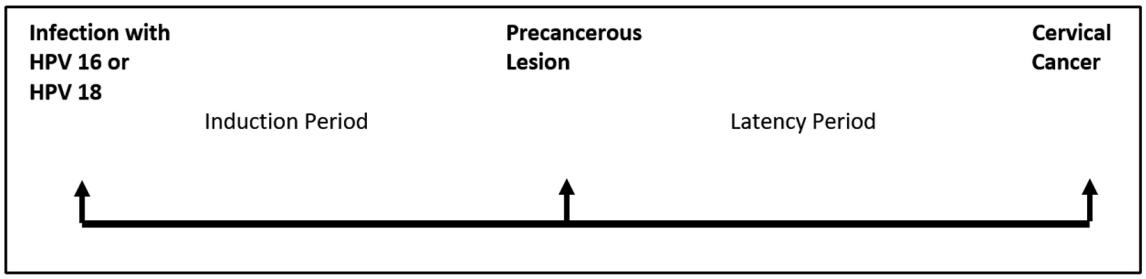


**Figure 2.** Induction period vs. latency period.

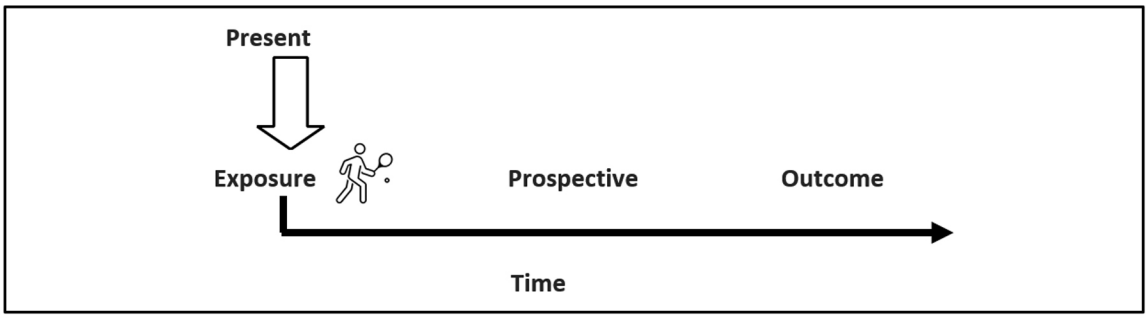


**Figure 3.** Prospective cohort study.

*Retrospective Cohort Studies*

An alternative approach is a retrospective cohort study, also known as a historical cohort study, wherein all the events under investigation transpired prior to the commencement of the study. In this cohort study, historical records are usually investigated to determine disease outcome and exposure status (refer to Figure 4).[3,5-7] An example is a researcher who launched his study in 2023 then went back to archived records in 1990 to define exposure then followed-up records until 2023 to determine outcomes.

Advantages

1. On occasion, costs may be reduced.
2. Provides more rapid results.

Disadvantages

1. Like prospective cohort studies, they are not appropriate for studying rare outcomes.
2. Utilizing records that were not originally intended for the study may result in data of substandard quality.
3. When data was collected in the past, information regarding potential confounding factors is often unavailable.
4. The task of identifying a cohort that is suitably exposed and a comparison group that is suitable can be demanding.
5. Additionally, differential losses to follow-up may introduce bias into retrospective cohort studies.

*Ambispective Cohort Studies*

There are both prospective and retrospective elements to an ambispective cohort design. Ambi-directional research is considerably less commonly used in comparison to prospective or retrospective studies (refer to Figure 5). However, these studies exhibit conceptual consistency with both classifications and possess certain aspects of their advantages and disadvantages.[3-7]

Open and Close Cohort Study

Cohorts may be classified as either open (dynamic) or closed (fixed). A close cohort consists of a fixed membership. Subjects cannot be added to a cohort once it has been established through enrollment and subsequent follow-up has commenced. In the event of mortality or loss of follow-up, the number of participants may decline; however, no additional participants are enrolled. Thus, over time, close cohorts invariably diminish in size. An open cohort, on the other hand, is dynamic in the sense that members may be added or removed at any time. A specific locality's cancer registry is an example of an open cohort. When a patient is diagnosed with cancer, additional subjects are consistently introduced. Individuals may also depart from the cohort through physical relocation or demise.[3]

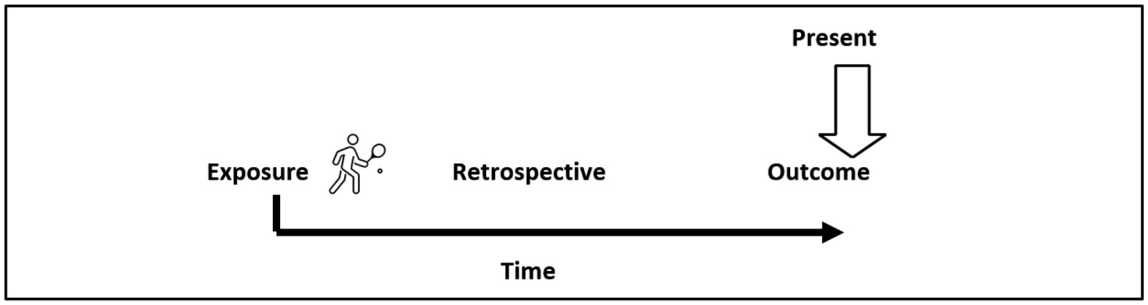


**Figure 4.** Retrospective cohort study.

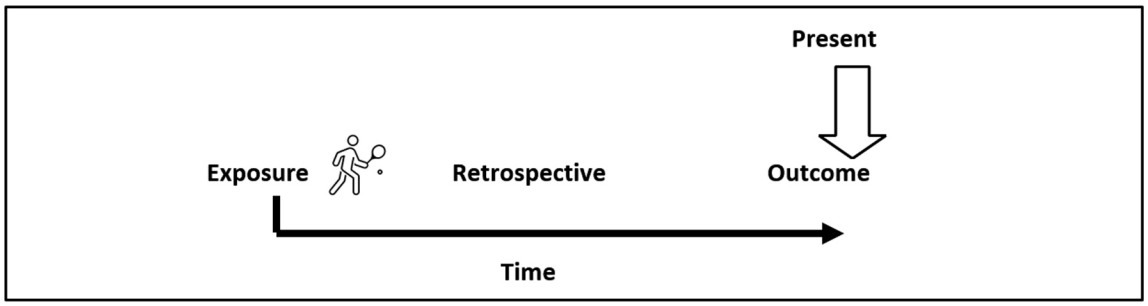


**Figure 5.** Ambispective cohort study.

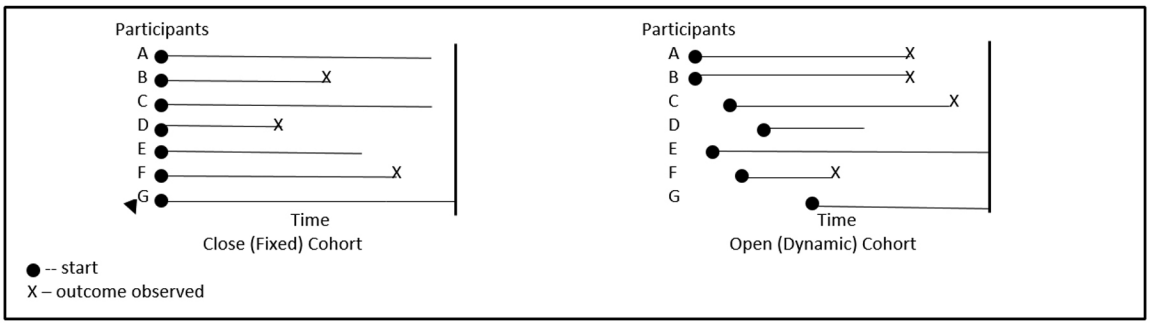


**Figure 6.** Close vs. open cohort.

Bias in Cohort Studies

*Bias due to confounding (Selection bias, Allocation bias, Case-mix bias, Channeling bias)*

Baseline confounding occurs when one or more prognostic variables (factors that forecast the outcome of interest) also predict the exposure received at baseline. Time-varying confounding occurs when individuals transition between exposures, and post-baseline prognostic factors influence exposure after baseline.[8-10]

*Bias in the selection of participants into the study (Selection Bias)*

When exclusion of some eligible participants, initial follow-up time for some participants, or some outcome events are related to both exposure and outcome, there will be a relationship between exposure and outcome even if the effects of the interventions are identical. Unlike confounding, this type of selection bias is distinct from confounding. A specific example of bias is the inclusion of frequent exposure users rather than new users.[8-10]

*Bias in the classification of interventions (Misclassification bias, Information bias, Recall bias, Measurement bias, Observer bias.)*

Differential or non-differential misclassification of exposure status introduces bias. Non-differential misclassification is unrelated to the outcome and tends to bias the estimated exposure effect toward the null. Differential misclassification occurs when misclassification of exposure status is associated with the outcome or risk of the outcome and is likely to result in bias.[8-10]

*Bias due to deviations from intended interventions (Performance bias; Time-varying confounding)*

Bias arises when systematic differences exist between exposure and comparator groups in the care provided, representing a deviation from the intended exposure(s). Assessment may be either the effect of assignment to intervention or the effect of starting and adhering to intervention.[8-10]

*Bias due to missing data (Attrition bias; Selection bias)*

The bias arises when later follow-up is missing for individuals initially included and followed (e.g., differential loss to follow-up due to prognostic factors); bias is due to the exclusion of individuals with missing information about intervention status or other variables, such as confounders.[8-10]

*Bias in the measurement of the outcome (Detection bias, Recall bias, Information bias, Misclassification bias, Observer bias, Measurement bias)*

Differential or non-differential errors in outcome data measurement cause bias. Such bias can occur when outcome assessors are aware of the exposure received, when different methods are used to assess outcomes in the groups, or when measurement errors are related to the status or effects of exposure.[8-10]

*Bias in the selection of the reported result (Outcome reporting bias; Analysis reporting Bias)*

Selective reporting of results in a way that depends on the findings.[8-10]

*Bias: "Healthy Worker's Effect"*

The "health worker" effect is a special selection bias in cohort studies of occupational exposures when the general population is used as the comparison group. The general population consists of both healthy people and unhealthy people. Those who are not healthy are less likely to be employed, while the employed workforce tends to have fewer sick people. Moreover, people with severe illnesses would be excluded from employment but not from the general population. As a result, comparisons of mortality rates between an employed group and the general population will be biased.[5,8-9]

Example of a Cohort Study in Family Practice

One example of a cohort study in family practice is the research by Mathews et al. on workers in the United States, which is published in the

Journal of Psychosomatic Research entitled Associations of job strain and family strain with risk of major depressive episode: A prospective cohort study in U.S. working men and women.[11]

**Steps in Conducting a Cohort Study**

Step 1*: Define the Research question and Objective(s)*

Finding a research question is the first step in the life cycle of a research question. One can identify problems or questions from patient care, teaching, reading, research, and practice organization. Eliciting issues or questions from patients, communities, colleagues, managers and networks, and finding existing data or research programs you can use, or join are other strategies one can use. Consider why your problem must be addressed and find its purpose since social value is essential in research. After finding a research question, one should refine it. This is done by reviewing related literature and answering three key questions: What is known? What knowledge gaps exist? What gap will this study fill? One can define the research question by answering the three key questions. Define each keyword. Consider using the PEO model to build a specific answerable question. Also, consider the need to identify secondary questions closely related to and aligned with the primary question. The interrogative statement(s) can be transformed into a declarative statement to become the research objective(s).[12] In the example cohort study, the objective was to provide evidence on sex differences in the relationship between job strain or family strain and risk of major depressive episode (MDE) within 12 months before the follow-up in United States (U.S.) employed men and women. The population is U.S.-employed men and women; the exposure is sex, job strain, and family strain. The observed outcome is a major depressive episode.

Step 2*: Identify the Study Population*

The second stage is population identification. The study population should be precisely defined and comprise all individuals who satisfy the study's eligibility criteria. The selected participant should not have the outcome(s) of interest. The cohort in the example consists of male and female employees from the MIDUS II cohort (baseline), and participants diagnosed with MDE within the past 12 months were excluded to minimize reverse causation. The research evaluated MDE using the 19-item Composite International Diagnostic Interview Short Form (CIDI-SF), a validated scale with high specificity and sensitivity. A diagnosis of a major depressive episode requires either a depressive mood or anhedonia for most of the day, as well as four or more symptoms (such as fatigue, appetite change, insomnia) for at least two weeks.

Step 3: *Determine the Sample Size*

Once the study population has been identified, the next step is determining the sample size. The sample size should be large enough to provide sufficient statistical power to detect meaningful differences between the exposure and outcome of interest. One can use Epi Info 7 to compute the sample size. Calculate the sample size recommended for a study given a set of parameters and the desired confidence level. The following are the steps to calculate a sample size for a cohort or cross-sectional study from Epi Info 7 (CDC, 2016).[13] Using three different statistical calculations, the application will show three different sample size estimates.

1. From the Epi Info™ main page, select StatCalc.
2. Select Cohort or cross-sectional. The Cohort or Cross-Sectional window opens.
3. Select the Two-sided confidence level of 95% from the drop-down list.
4. Enter the desired Power (80%) to detect a group difference at that confidence level.
5. Enter the proportion of unexposed vs. exposed. This is a single value; the approximate proportion cannot be entered in the format # of Unexposed: # of Exposed.
6. Enter the percentage outcome in the unexposed group. This percentage represents the incidence rate of ill patients in the unexposed group.
7. Enter the percentage outcome in the exposed group. This percentage represents the incidence rate of ill patients in the exposed group.
8. The Risk ratio and Odds ratio fields automatically populate based on the values entered.
9. The output table shows three different sample size estimates.

Step 4: *Define and Select the Exposure Groups*

*Selection of subjects from the cohort who were exposed*

In a cohort study, the exposed group is selected based on the hypothesis being tested, the frequency of exposure, and practical considerations such as the availability of records and the ease of follow-up. Special exposure cohorts investigate the health effects of uncommon exposures or risk factors, such as uncommon workplace contaminants, unusual cuisines, and unusual lifestyles. Frequently, strata are selected from occupational groups (such as coal miners) or religious groups (such as Islam) where known exposures occur. The general population cohort is used for prevalent exposures or risk factors, such as smoking and alcohol consumption. To facilitate precise follow-up and determination of the investigated outcome, these cohorts are chosen from professional groups, such as nurses, or well-defined geographic areas. After enrolling a general population cohort, researchers will ascertain their baseline exposures to many exposures of interest and potential confounding factors that may require adjustment in the analyses.[4] The general population cohort of workers in the U.S. was used in the sample cohort. The baseline exposure measurements included the participants' sex, the presence of job strain, and the presence of family strain. Job strain was defined using Karasek's Job Demand-Control model, whereas family strain was measured using a four-item questionnaire regarding familial stressors. Other variables such as age, race, marital status, education, annual household income, current smoking, alcohol consumption, and frequency of vigorous recreational physical activity were also collected.

*Choosing unexposed or controls*

In a cohort study, there are three potential comparison groups: an internal comparison group, the general population, and a comparison cohort. An internal comparison group is comprised of unexposed cohort members. In general, this is the best comparison group, as the subjects are comparable in numerous ways. In the cohort used as an example, an internal comparison group was used. When no comparable internal comparison group is available, the general population is used for comparison. The comparison with the general population is based on preexisting population data on disease incidence or mortality, such as atypical occupational exposure. However, the general population may differ from the exposed workforce in a variety of ways, including in terms of overall health.[4] A comparison cohort is comprised of individuals from another cohort. It is the least preferable option because the comparison cohort, despite not being exposed to the exposure under study, is frequently exposed to other potentially hazardous substances, making it difficult to interpret the results.

Step 5: *Follow up the study population and observe the appearance of the outcome.*

The study population should be followed up over a defined period to determine whether the outcome of interest develops. Measurement of the outcome during follow-up can be conducted through various methods, such as a review of medical records or questionnaires.[4] In the example cohort, the tool that was used to exclude participants for the presence of MDE was used during follow-up (MIDUS III) to ascertain the outcome.

Step 7: *Analyze the Data.*

Once the data has been collected, it should be analyzed to determine the relationship between the exposure and outcome of interest. Various statistical methods, such as t-test, chi-square test, or regression, can be used to analyze the data.[3]

*T-tests*

In epidemiology, it is common to have two samples representing two distinct populations, with one sample answering queries about whether the population means of the two populations are sufficiently different to conclude that the populations they represent have different means. Under the null hypothesis, the t-test employs a statistic to determine whether the two means differ significantly.[3]

*Chi-squared test*

Cross tabulations, also known as contingency tables, are tools for displaying a number of participants classified by two or more factors or variables. A 2 x 2 table is an example of a contingency table in which the association between two exposure (exposed or unexposed) and two outcome status categories (with disease or without disease) is displayed. Close examination of the table inevitably raises the question of whether there is evidence of an association between exposure and disease.[3]

*Regression*

Regression models are indispensable for data analysis and are widely employed in epidemiological research. Although their underlying concepts are simple, the calculations can be complicated. Thankfully, computer programs can perform the necessary calculations. As such complexity is unnecessary for this text, we will concentrate on applying and interpreting these methods.[3]

*Types of regression models[3]*

1. Linear Regression – the dependent variable needs to be continuous with its frequency distribution being normal.

2. Logistic regression – the dependent variable is derived from the presence or absence of a characteristic, typically represented as 0 or 1.

3. Cox proportional hazard regression, a type of survival analysis – the dependent variable represents the time from baseline for some kind to the occurrence of an event of interest.

*Survival Analysis*

Done in Cox proportional hazard models – has an additional complexity in that censoring status must also be considered.[3] Survival Data or Time-to-event data are measurements of elapsed time between the initial enrollment in a study and the final disposition (outcome) of the study subject. This elapsed time could be represented by the time of initial diagnosis, or it could be represented by the point in time when one enters the study. Survival in this context simply means that an event has not occurred, not, necessarily, that the endpoint of interest involved an examination of "life" and "death".[14]

Step 8: *Write the Final Report*

**Ethical issues**

The National Ethical Guidelines for Research for Involving Human Participants (2022) details the key ethical issues in conducting epidemiological research.[15]

Key issues that were mentioned are:

1. The collection of data from individuals who may not directly benefit from prospective public health interventions and who frequently do not require treatment for a disease. Consequently, it is essential to ensure that the research risk is minimal, and the societal benefit is beneficial.

The STROBE guideline

| | Recommendation |
|---|---|
| **Title and abstract** | (*a*) Indicate the study's design with a commonly used term in the title or the abstract |
| | (*b*) Provide in the abstract an informative and balanced summary of what was done and what was found |
| **Introduction** | |
| Background/rationale | Explain the scientific background and rationale for the investigation being reported. |
| Objectives | State-specific objectives, including any prespecified hypotheses |
| **Methods** | |
| Study design | Present key elements of study design early in the paper |
| Participants | (*a*) Give the eligibility criteria and the sources and methods of selection of participants. Describe methods of follow-up |
| | (*b*) For matched studies, give matching criteria and the number of exposed and unexposed |
| Variables | Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable. |
| Data sources/ measurement | For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe the comparability of assessment methods if there is more than one group |
| Bias | Describe any efforts to address potential sources of bias |
| Study size | Explain how the study size was arrived at |
| Quantitative variables | Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why |
| Statistical methods | (*a*) Describe all statistical methods, including those used to control for confounding |
| | (*b*) Describe any methods used to examine subgroups and interactions |
| | (*c*) Explain how missing data were addressed |
| | (*d*) If applicable, explain how loss to follow-up was addressed |
| | (*e*) Describe any sensitivity analyses |
| **Results** | |
| Participants | (a) Report numbers of individuals at each stage of study—e.g. numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analyzed |
| | (b) Give reasons for non-participation at each stage |
| | (c) Consider the use of a flow diagram |
| Descriptive data | (a) Give characteristics of study participants (e.g., demographic, clinical, social) and information on exposures and potential confounders |
| | (b) Indicate the number of participants with missing data for each variable of interest |
| | (c) Summarize follow-up time (e.g., average and total amount) |
| Outcome data | Report numbers of outcome events or summary measures over time |
| Main results | (*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included |
| | (*b*) Report category boundaries when continuous variables were categorized |
| | (*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful period |
| Other analyses | Report other analyses are done—e.g., analyses of subgroups and interactions and sensitivity analyses. |
| **Discussion** | |
| Key results | Summarize key results with reference to study objectives. |
| Limitations | Discuss the study's limitations, considering sources of potential bias or imprecision. Discuss both the direction and magnitude of any potential bias. |
| Interpretation | Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence. |
| Generalizability | Discuss the generalizability (external validity) of the study results. |
| **Other information** | |
| Funding | Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based |

2. Even though epidemiologic research typically does not involve interventions that may cause physical distress to eligible participants, these studies still require their time and attention. They might violate the right to privacy and confidentiality. Psychological harms such as embarrassment, intense emotional reactions, and social hazards must be considered.
3. In observational or non-interventional epidemiologic studies, consent procedures do not need to be as rigorous as in clinical trials of novel drugs and treatment modalities. However, the Research Ethics Committee (REC) scrutinizes the protocol and determines whether such non-disclosure is justified when the researcher proposes selective disclosure of information (e.g., "blinding").
4. Genetic and other biological materials are frequently collected in epidemiologic studies. The RECs and other relevant authorities are responsible for establishing the conditions for using these materials beyond their epidemiologic objectives. (See section on Human Data and Samples Obtained from Biobanks, Registries, and Databases Research)
5. There are conflicts of interest in epidemiologic studies, though they may not be as evident as in intervention research such as clinical trials. Financial interests and a researcher's ideologies may influence scientific judgment and study outcomes. For instance, the marketing of vaccines in developing nations may be based on the prevalence of a disease as determined by an epidemiological study or public health program and may be influenced by epidemiology data prompted by advocacy.

## References


1. Aschengrau A & Seage GR. Essentials of Epidemiology in Public Health. Jones & Bartlett 2013.
2. Ben-Shlomo Y, Brookes ST & Hickman M. Epidemiology, evidence-based medicine and public health ecture Notes 1st ed. Wiley-Blackwell. 2013
3. Bonita R, Beaglehole R & Kjellstrom Y. Basic Epidemiology 2nd ed. World Health Organization. 2006
4. Greenberg RS & Lackland DT. Cohort studies. Greenberg R.S.(Ed.), Medical Epidemiology: Population Health and Effective Health Care, 5e. McGraw Hill. 2014 https://accessmedicine-mhmedical-com.feunrmf.remotexs.co/content.aspx?bookid=1430&sectionid=84052073
5. Boston University. Cohort Studies. 2016 https://sphweb.bumc.bu.edu/otlt/MPH-Modules/EP/EP713_CohortStudies/
6. Euser AM, Zoccali C, Jager KJ & Dekker FW. Cohort studies: prospective versus retrospective. Nephron Clin Pract 2009; 113(3): c214–c217. https://doi.org/10.1159/000235241
7. Hammill BG. Chapter 12. observational study designs. Lopes R.D., & Harrington R.A.(Eds.), Understanding Clinical Research. McGraw Hill. 2013 https://accessmedicine-mhmedical-com.feunrmf.remotexs.co/content. aspx?bookid=674&sectionid=45407263
8. Ramirez-Santana M. Limitation and biases of cohort studies. Intech Open. Creative Commons Attributions NonCCOmmercials. 2016 http://dx.doi.org/10.5772/intechopen.74324
9. Davies HT & Crombie IK. Bias in cohort studies. Hospital medicine (London, England: 1998) 2000; 61(2): 133–5. https://doi.org/10.12968/hosp.2000.61.2.1280
10. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.4 (updated August 2023). Cochrane, 2023. Available from www.training.cochrane.org/handbook.
11. Matthews TA, Robbins W, Preisig M, von Känel R, Li J. Associations of job strain and family strain with risk of major depressive episode: A prospective cohort study in U.S. working men and women. J Psychosom Res 2021 Aug;147:110541. doi: 10.1016/j.jpsychores.2021.110541. Epub 2021 Jun 8.
12. Goodyear-Smith F & Mash B. How to Do Primary Care Research. CRC Press, Taylor & Francis Group. 2019
13. Center for Disease Control. Epi Info 7 User Guide. Version 3. 2016 https://www.cdc.gov/epiinfo/pdfs/UserGuide/EI7Full.pdf
14. Daniel WW & Cross LC. Biostatistics: A Foundation for Analysis in the Health Sciences, 10th Edition: A Foundation for Analysis in the Health Sciences. Wiley. 2012
15. Philippine Health Research Ethics Board. National Ethical Guidelines for Research Involving Human Participants. Philippine Council for Health Research and Development. 2022