ERIC Notebook

Second Edition

Cohort Studies

Second Edition Authors:

Lorraine K. Alexander, DrPH

Brettania Lopes, MPH

Kristen Ricchetti-Masterson, MSPH

Karin B. Yeatts, PhD, MS

A cohort study is a type of epidemiological study in which a group of people with a common characteristic is followed over time to find how many reach a certain health outcome of interest (disease, condition, event, death, or a change in health status or behavior). A cohort is defined as a group of persons, usually 100 or more in size, who share a common characteristic, e.g. smokers, workers in a lead smelter, people born in the same year, or all enrollees of a specific health insurance plan. Cohort studies compare an exposed group of individuals to an unexposed (or less exposed) group of individuals to determine if the outcome of interest is associated with exposure. There are two types of cohort studies: prospective and retrospective (or historical) cohorts. Prospective studies follow a cohort into the future for a health outcome, while retrospective studies trace the cohort back in time for exposure information after the outcome has occurred. Both types of cohort studies are also referred to as longitudinal or follow-up studies.

Establishing the cohort

The investigator controls the selection of the cohort. The investigator may choose a cohort based on age, location, exposure to a certain working environment, or some other common characteristic. Cohorts may be selected on the basis of exposures known at baseline, e.g. smokers vs. nonsmokers. Alternatively, cohorts may be divided into exposure categories once baseline measurements of a defined population are made. For example, the Framingham Cardiovascular Disease Study (CVD) used baseline measurements to divide the population into categories of CVD risk factors.

For instance, an investigator wants to study whether exposure to military aircraft engine noise is a risk factor for hearing loss. The cohort this investigator would want to establish should be composed of two groups of military personnel: one exposed to engine aircraft noise (the group under study) and the other unexposed to engine aircraft noise (a comparison group). The unexposed group should be representative of the exposed group on all factors except exposure.

The cohort at baseline

After the cohort of study subjects is established, their individual exposures of interest are identified at baseline (through interviews, questionnaires, bioassays, medical records, etc.). Subjects with the outcome of interest at baseline are excluded. Therefore, all members of the cohort are at risk of developing the outcome at the beginning of observation.

Following the last example, anyone in the cohort of military personnel with a specified hearing loss at baseline would be excluded from the cohort and would not be followed.

Following the cohort

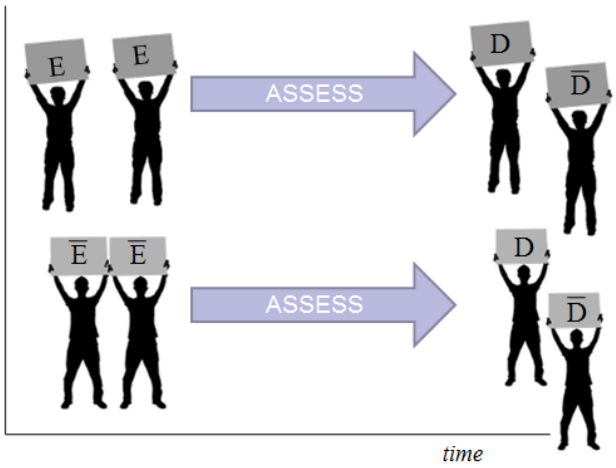
The cohort is then followed over time for new occurrences of the outcome of

UNC GILLINGS SCHOOL OF GLOBAL PUBLIC HEALTH

interest, in the above example, hearing loss. In a prospective, or concurrent, cohort study baseline exposure is assessed at the beginning of the study and the cohort is followed into the future. In a retrospective, or historical cohort study, baseline exposure is assessed at some point in the past through historical records, e.g. health records for a cohort of factory workers may provide exposure and outcome information up to the present.

Cohort	Baseline	Follow-up
Prospective	Assessed at beginning of study	Followed into the future for out-
Retrospective	Assessed at some point in the past via historical records	Outcome has already occurred and is assessed via historical rec-

Cohorts are followed over time to the end of follow-up. Occurrence of the outcome of interest may be determined via interviews with members of the cohort and/or family members, or by viewing health and/or work records to conclude the study.



The basic design of a cohort study from beginning of the study to end of follow-up. E = exposed, Ē = not exposed, D = diseased and D̄ = not diseased.

Evaluation of the results

During the follow-up period the investigator counts the number of subjects who develop the outcome of interest. This count is the numerator for a calculation of risk, also referred to as **cumulative incidence** or **incidence proportion**. The number of persons at risk at baseline is the denominator.

Risk (also called Cumulative Incidence or Incidence Proportion) = new occurrences of the outcome / population-at-risk at baseline

Two risks can be compared to provide a risk ratio. The reference group is a comparable unexposed cohort. The index group is the exposed cohort. The risk ratio is computed by dividing the risk in the exposed group by the risk in the unexposed group. The risk ratio gives a relative measure of the increase or decrease in incidence between the exposed and unexposed groups.

$$\text{Risk Ratio} = Risk_{exposed}/Risk_{unexposed}$$

As with risk, **an incidence rate measure (IR)** is calculated with new occurrences of the outcome as the numerator. An incidence rate is also called incidence density. However, in an IR calculation the denominator is person-time (days, months, or years) at risk during follow-up. Person-time is measured by summing the total time each member of the cohort was free of the outcome of interest and thus contributed to person-time-at-risk during the follow-up period. The IR measures the rapidity of occurrence of new health outcomes in the population.

$$\text{Incidence rate} = \frac{new\ occurrences\ of\ the\ outcome}{person\text{-}time\ at\ risk}$$

Two IRs may also be compared to find the relative increase or decrease in the rate of health outcome occurrence between the exposed and unexposed groups. This relative measure is called the **incidence rate ratio (IRR) or the rate ratio.**

$$\text{Incidence rate ratio} = \frac{IR_{exposed}}{IR_{unexposed}}$$

Incidence measures between exposed and unexposed cohorts can also be subtracted from one another to find the difference between the two

measures. This measure is referred to as a **risk difference** or a **rate difference**.

$$\text{Risk Difference} = Risk_{exposed} - Risk_{unexposed}$$

$$\text{Rate Difference} = I_{exposed} - I_{unexposed}$$

Exposure may be a risk factor or a preventive factor in the development of the outcome of interest. When exposure is preventive, the risk ratio or rate ratio will be less than one.

Advantages of a cohort study

A cohort study can be used to directly measure the risk and rate of a health outcome occurrence over time. Cohort studies are an efficient means of studying rare exposures (e.g. gasoline fumes, as discussed in the next paragraph), in contrast to case-control studies, which tend to be better for rare outcomes. Cohort studies also allow the investigator to assess multiple outcomes of a single exposure.

> A cohort study would be the most efficient means of studying the effects of long-term exposure to gasoline fumes. The cohort would consist of individuals who are exposed daily to gasoline fumes (auto mechanics, gas station attendants, sea crewman on tankers, etc.). By studying this group
> of individuals, the investigator can better determine the direct effects of long-term, regular gasoline inhalation. Also, by conducting a cohort study, an investigator could determine if gasoline inhalation causes many different health outcomes (e.g., different types of cancer and respiratory illnesses).

Additional advantages of cohort studies

Cohort studies establish temporal relationships between exposure and outcome. Exposure clearly precedes the outcome because the population under study at baseline is free of the outcome of interest. Cohort studies also avoid recall bias (as the exposure is determined before the outcome, one's health outcome or disease state won't affect how accurately one recalls exposure levels), as well as, survival bias (duration of disease influencing exposure measurements). Therefore, cohort studies are the best observational study design used to help establish cause and effect relationships.

Disadvantages of cohort studies

Cohort studies often require large sample sizes, especially when the outcome is rare, defined as less than 1 event per 1000 person- years (e.g., all specific cancers). Therefore, cohort studies tend to be expensive and time-consuming. When there are *losses to follow-up* (individuals who leave the cohort before the end of follow-up) biases may occur. Thus, individuals who leave the cohort prematurely may have a different baseline risk than the members who remain in the cohort throughout the entire length of follow-up. Therefore, the study may not be *generalizable* to the original target population, but only to those who remained under investigation throughout the length of the study. Also, any differences in the quality of measurement of exposure or disease between exposed and non-exposed cohorts may introduce *information bias* and thereby distort the results.

References

Dr. Carl M. Shy, Epidemiology 160/600 Introduction to Epidemiology for Public Health course lectures, 1994-2001, The University of North Carolina at Chapel Hill, Department of Epidemiology

Rothman KJ, Greenland S. Modern Epidemiology. Second Edition. Philadelphia: Lippincott Williams and Wilkins, 1998.

The University of North Carolina at Chapel Hill, Department of Epidemiology Courses: Epidemiology 710, Fundamentals of Epidemiology course lectures, 2009-2013, and Epidemiology 718, Epidemiologic Analysis of Binary Data course lectures, 2009-2013.

Acknowledgement

The authors of the Second Edition of the ERIC Notebook would like to acknowledge the authors of the ERIC Notebook, First Edition: Michel Ibrahim, MD, PhD, Lorraine Alexander, DrPH, Carl Shy, MD, DrPH and Sherry Farr, GRA, Department of Epidemiology at the University of North Carolina at Chapel Hill. The First Edition of the ERIC Notebook was produced by the Educational Arm of the Epidemiologic Research and Information Center at Durham, NC. The funding for the ERIC Notebook First Edition was provided by the Department of Veterans Affairs (DVA), Veterans Health Administration (VHA), Cooperative Studies Program (CSP) to promote the strategic growth of the epidemiologic capacity of the DVA.