# **Lecture 3: Ecologic & Cross-sectional studies**

# **Learning objectives**

By the end of this session, participants should be able to:

1. Describe the basic design features of ecologic studies and cross-sectional studies.
2. Decide when it would and would not be appropriate to use these study designs.
3. Appreciate the strengths and weaknesses of these study designs.

# **1. Ecologic studies**

Ecologic studies are ones in which *a measure of exposure is assigned to* ***a population group*** (as opposed to individual persons) on the basis of time and place. For example, air pollution monitoring stations may be used to classify the exposure to outdoor air pollutants of people within different towns and cities, which is then related to a health outcome, for example the incidence of asthma. Or survey data on malaria prevalence in different regions may be used to examine the geographical relationship between malaria and the incidence of nephrotic syndrome. Measurements of both exposure and outcome, for example air pollution and malaria, are not made in individuals but are made at a population level.

Therefore, in contrast to the other analytic epidemiologic study designs where we are concerned with these associations in individuals (analytic cross-sectional, case-control and cohort studies), the unit of analysis for an ecologic study is an entire population. Ecologic studies often make use of routinely (e.g. health facility records, sentinel surveillance) or frequently (e.g. Demographic and Health Surveys (DHS), National Health Surveys) collected sources of data. In all ecological studies it is critical that the exposure and outcome have been measured consistently and in the same way over time and place.

## **1.1. Geographic Studies**

The most common type of ecologic study is a geographic study, such that we classify population groups by some characteristic of the area in which they live or work, and then look to see the relationship between this characteristic and health. For example, a study might obtain information on the average intake of salt in a population and the mortality rate from stroke in that same population. This information can be collected for several different countries and the data examined for a relationship between the salt intake of a population and rate of disease. Or you might examine life expectancy by country according to per capita income (see figure below, from [Gapminder](http://www.gapminder.org)). Often a scatter diagram is plotted, where each dot represents one region/country, and a regression line is drawn to estimate the strength of the relationship between the exposure and outcome at a group level (see Appendix I).

Geographical studies are often used in epidemiology because they are usually simple and inexpensive to undertake. Geographic studies are ideally suited to investigate associations on a population level, such as for the effects of widespread environmental exposures, laws/policies, or population characteristics without an individual analogue (e.g. income inequality, housing density, racial diversity).



## **1.2. Temporal Studies**

For temporal studies, the exposure and outcome are measured within a single population over multiple points of time. For example, one can measure population-level changes in mortality over time and compare these to changes in diet or housing density over time. In these studies, the time trend in population exposure is compared to the time trend in a disease outcome in the same population. The simplest form of this is a before-after comparison when an intervention has been applied. Since many other factors vary over time other than the one of interest, these comparisons are only helpful when one can exclude the possibility of other exposures being responsible for the difference in outcome.

An example of a temporal study is shown below. The data come from Japan around the time of the Tōhoku earthquake and tsunami in 2011.

The figure is a line graph showing the number of pneumonia deaths over time by three separate areas in Japan: “All municipalities”, “Inland municipalities”, and “Coastal municipalities”. The middle of the figure has a vertical line indicating the date of the Tōhoku earthquake and tsunami (11th March 2011). 

The horizontal axis has no title, and the tick values range from “Pre 52nd week” (indicating the 52nd week prior to the Tōhoku earthquake and tsunami) to “Post 49th week” in intervals of 4 weeks. 

The vertical axis has the title “No. of pneumonia deaths” and the tick values range from 0 to 250 in intervals of 50.

The figure broadly shows that the number of pneumonia deaths was 25-50 in coastal municipalities and 50-100 in inland municipalities in the year preceding the earthquake, increased suddenly following the earthquake to more than double, then returned to numbers similar to before the earthquake approximately three months later.

Figure 2. Shibata Y et al. “Characteristics of pneumonia deaths after an earthquake and tsunami: an ecological study of 5.7 million participants in 131 municipalities, Japan” BMJ Open. 2016 Feb 23;6(2).

A second issue with temporal ecological studies is that there may be a considerable (or unknown) lag time between the exposure and disease. For example, the policy in an area may change to promote uptake of the hepatitis B vaccine. The impact on the incidence of liver cancer can take decades to detect. In this situation you do not know which time points for exposure should be compared to which time points for disease.

## **1.3. Ecologic fallacy**

The quantitative relationship between exposure and disease can be obtained from the regression of disease frequency against prevalence of exposure. However, the results of analyses based on grouped data apply at grouped level, and not necessarily at individual level. To assume that grouped results apply to individuals is to make the **ecologic fallacy.**

|  |  |
| --- | --- |
|  |  |
| ① Data pattern/relationships at the group level also applies at an individual level. Hence, ecologic fallacy not present.  ② Data pattern/relation at the group level is not applicable at the individual level. | |

A famous example by Durkheim is the correlation of suicide rates with religion in regions of Prussia. Durkheim found a positive association between the proportion of Protestants in a region and the rate of suicide. This could be interpreted as evidence that Protestants were more likely than Catholics to die from suicide (perhaps because of their weaker social support structures); but another possibility is that the high suicide rate in predominantly Protestant regions was due to suicides in Catholics in these regions who were socially isolated. Without data at individual level, it is unclear which if either of these explanations represents cause and effect: the group classification of majority religion is not a good measure of religion of the individual, and it is not therefore possible to draw robust conclusions about relationships at the individual level.

### **1.3.1. Advantages and disadvantages of ecologic studies**

**Advantages:**

* Often relatively cheap and quick to carry out
* Appropriate for analysis of population-level features (e.g. per capita income, laws, low emission zones)
* Uses routine or frequently collected data (e.g. routine facility records, Demographic Health Surveys (DHS), [WHO’s World Health Statistics](https://www.who.int/data/gho/data/themes/world-health-statistics)/Global Health Observatory). See more in Appendix II - Routine Data 2019.docx.
* Differences in exposure between areas or over time may be larger than those between individuals
* Ecologic studies are useful for generating hypotheses to be tested at individual level

**Disadvantages:**

* Data are often unavailable on confounding factors
* Measures of exposure are only surrogates based on the average in the population: grouped results cannot reliably be extrapolated to the individual level (e.g. per capita income)
* There may be systematic differences between areas or over time in recording of disease frequency:

- quality of diagnosis

- disease classification

- completeness of reporting

- differential survival

- population under/over enumeration

* There may be systematic differences in the measurement of exposures
* Spatial boundaries sometimes artificially divide populations in ways that may obscure the true distribution of exposure and disease risk

## **1.4. Ecologic studies are sometimes the design of choice**

Ecologic studies are not always the “poor relations” in the family of epidemiological designs. Sometimes they are the appropriate choice. This is when one is specifically interested in determinants of the health of populations per se rather than when you are using them as an efficient proxy for looking at an association that you would ideally study at the level of individuals. Ecologic studies are the design of choice when for example one is investigating the influence of population-level characteristics (e.g. national wealth) on variation in disease rates or mortality between populations. This is exemplified by the classic study by Samuel Preston of GNP per capita and life expectancy at birth for national populations published first in 1975 and then republished in the International Journal of Epidemiology in 2007 with commentaries (Preston SH. The changing relation between mortality and level of economic development. Population Studies, Vol. 29, No. 2, July 1975. Int J Epidemiol 2007; 36(3): 484-90).

# **2. Cross-sectional Studies**

Cross-sectional studies are a simple type of epidemiological study. Information about health-related measures is collected from each individual *at one point in time*.

Information can be gathered about outcomes (e.g. diseases, infections, or other conditions of interest), about exposures (e.g. smoking, diet), or about both outcomes and exposures at the same time. When you see the word “survey” in the literature or the press, it usually means a cross-sectional study (but not always, which is why we need to be more precise with our terminology!).

Most cross-sectional studies are done on *samples* of the population rather than the whole population (e.g. a census, which is ideal, but usually impractical). The way in which the sample is drawn from the study population is critical to the **representativeness** of the sample in relation to the study population. Although there are many alternative ways to draw a sample from a population, a general rule is that some variation of a **random sample** is ideal.

The mainmeasure of outcome frequency obtained from a cross-sectional study is **prevalence**. As we saw in Lecture 1, prevalence measures the proportion of the population who have the outcome (or exposure) of interest. It is calculated as the number of people with the outcome in a population at one point in time, divided by the total number of people in that population at that time. Since prevalence is a proportion, it cannot be less than zero or greater than 1.0 (or 0%-100%).

Alternatively, one can calculate a prevalence odds:

You will notice we used the word “case” in the above equations. “Case” is often used to refer to the people with the outcome of interest in a cross-sectional study, e.g. those who are HIV positive. It is fine to use this term, but be aware that (i) being a “case” does not necessarily mean that the person has a disease (e.g. “cases” might be those who are currently pregnant in a population) and (ii) that cross-sectional studies are not the same as case-control studies, and people in a cross-sectional study who are not cases are referred to as non-cases (not “controls”). You will learn more about case-control studies and how they differ from cross-sectional studies in a later session (see lecture on case-control studies).

Prevalence can be thought of as the frequency of the outcome in a population at a *point in time* and that is why it is more correctly called **point prevalence**. However, point prevalence is often simply referred to as **prevalence.**

## **Example:**

A cross-sectional survey is undertaken in Kenya. 5000 people aged 20-50 years are examined. Of these, 200 people have diabetes and 4800 people do not. The point prevalence of diabetes in this population is therefore:

An alternative calculation is the prevalence odds of diabetes, which is:

We can also use cross-sectional studies to measure the *prevalence of exposure*. For example, in this same cross-sectional survey in Kenya, we can measure the prevalence of smoking, low socio-economic status, adverse health behaviours, and so on.

Questions which relate to the prevalence of a characteristic *at any point during a period of time in the past* provide an estimate of the **period prevalence**. An example of a question that measures period prevalence is ‘Have you had diarrhoea on any day *within the past week*?’. In contrast, an example of a question that measures point prevalence is: ‘Do you have diarrhoea *today*?’. The period prevalence will include episodes of diarrhoea which started (were incident) before the start of the past week (the reference period) but continued into it, *and* episodes which started during the reference period.

Most cross-sectional studies include a mix of questions and/or measurements on the current status of the study participant (e.g., current symptoms or conditions, current blood pressure) and questions about the past (e.g., a child's age when breastfeeding was stopped). These questions still form part of a cross-sectional study, because the information is collected at one point in time. Inclusion of questions about the past (often called retrospective questions) enable epidemiologists to answer analytic questions pertaining to the relationships between different disease determinants and health outcomes (more discussion in 2.1.2).

As mentioned, cross-sectional studies are usually conducted in a representative sample of the population of interest. The prevalence of the outcome or exposure obtained from in the sample is assumed to be applicable to the population of interest as well. Prevalence therefore provides a measure of the burden of disease, or the prevalence of specific exposures, in the population of interest. Using such information, public health planners and administrators can know what services are required to respond to needs in the population, and determine the allocation of health care resources in the community accordingly. Repeated cross-sectional studies over a period time are also useful for monitoring disease trends or to monitor interventions to see if there is an impact on the prevalence of disease (or exposure). Cross-sectional studies can often be performed more quickly and economically than other types of study design, making them a useful and efficient tool for public health.

## **2.1. Purposes of cross-sectional studies**

### **2.1.1. Descriptive:**

Cross-sectional studies allow the collection, at one point in time, of information on the frequency and distribution of health-related variables in the study population. These are often referred to as prevalence surveys.

As discussed above, the frequency of a disease, condition, or characteristic in a population at one point in time is known as the **point prevalence.**

The point prevalence of angina among adults in Mongolia would be an example of a result from this kind of study. These angina prevalence data could be broken down by age, sex and geographical area, for example, to give a more detailed picture of which groups are particularly affected. If the data have been collected on a representative sample of all adults in Mongolia, they will have the advantage over routine health service statistics on angina since a representative sample includes adults who have not made contact with the health services.

### **2.1.2. Analytic:**

In an analytic cross-sectional study, information about both exposure(s) and outcome(s) are measured from individuals, often with an aim to determine if there is a causal relationship between the two, or if the distribution of exposure or outcome is inequitable across groups.

Taking one exposure or potential risk factor at a time, the prevalence of disease is compared between the exposed (with the risk factor) group and the unexposed (without the risk factor) group. This comparison is done by dividing the prevalence of disease in the exposed group by the prevalence of disease in the unexposed group to calculate the **prevalence ratio**.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **With disease** | **No disease** | **Total** |
| Exposed | A | b | a+b |
| Not exposed | c | d | c+d |
| TOTAL | a+c | b+d | a+b+c+d |

For example, the prevalence of tobacco smoking in a population in 2005 was found to be 33.1% for men (15 years or older) and 3.8% for women (15 years or older). The prevalence ratio for tobacco smoking is:

This shows that in 2005, men had 8.7 times the prevalence of tobacco smoking than women. As an alternative, it is also possible to calculate a prevalence odds ratio:

This shows that in 2005, men had 12.5 times the prevalence odds of tobacco smoking compared to women.

And it is possible to calculate a prevalence difference:

Which is to say that the prevalence of smoking among men was 33.1% and the prevalence of smoking among women was 3.8%, a difference of 29.3%. For exposure-outcome associations which are shown to be causal (see lecture on Confounding & Causation), the prevalence difference can be interpreted as an attributable risk (see lecture on Epidemiologic Measures).

One of the key features of analytic cross-sectional studies is that the information on the potential risk factors and outcomes are measured simultaneously. Cross-sectional studies can be repeated in the same population over time, but each round will recruit a different random sample of people. It is important to remember that in cross-sectional studies, individuals are not followed over time.

The fact that cross-sectional studies are limited to the measurement of risk factors and outcomes at one, simultaneous, point in time imposes a limitation. It is often difficult to measure, or for people to remember, exposures which happened a long time before the survey. These might be the exposures that are causally related to the disease. For example, hepatitis B infection is causally related to clinical hepatocellular carcinoma but the mean interval between hepatitis B and onset of the cancer is about 40 years. And the interval between HIV infection and the development of AIDS is often 10 years or more.

Exposures which are the most well-suited to being investigated by cross-sectional studies as potential risk factors for diseases are those which do not change much over time. Genetic characteristics or other static traits such as blood group are good examples of these factors. Exposures which are less well-suited to assessment as potential risk factors in cross-sectional studies are those which do change over time, such as diet, especially since diet may change due to having a particular health outcome, rather than being a risk factor for the outcome.

## **2.2. Methodological issues for cross-sectional studies**

* Measurement bias (see lecture on Bias and Error)
* Selection bias (see lecture on Bias and Error)
* Sample size (see lecture on Bias and Error)
* Confounding (see lecture on Confounding & Causality)
* Temporality (see lecture on Confounding & Causality)

# **Further reading**

Bailey L, Vardulaki K, Langham J and Chandramohan D, *Introduction to Epidemiology*. Chapter 4 and 5. Open University Press, 2005 (Understanding Public Health, Series editors: Nick Black and Rosalind Raine)

Dos Santos Silva, I. *Cancer Epidemiology: Principles and Methods*, Chapter 10 and 11. IARC, Lyon, France. 1999

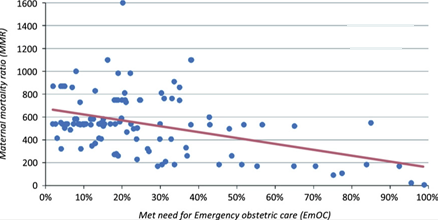
Greenland S, Morgenstern H. Ecological bias, confounding and effect modification. *Int J Epidemiol* 1989; **18:**269-74

Hennekens CH & Buring JE, *Epidemiology in Medicine*, Chapter 5. Little, Brown and Company, 1987.

Webb P and Bain C. *Essential Epidemiology: An introduction for Students and Health Professional*s. Chapter 2. Second Edition. Cambridge University Press. 2011.

**Appendix I**

In statistics, a *regression line* is a line that describes the way in which an exposure and an outcome are related (e.g., as one goes up so does the other, or as one goes up the other goes down. An example is given below -



Examples of using a regression line to describe/represent the relationship between an exposure and an outcome in an ecological study. Each circle represents a country-year. (Source: Holmer, H., Oyerinde, K., Meara, J.G., Gillies, R., Liljestrand, J. and Hagander, L., 2015. The global met need for emergency obstetric care: a systematic review. BJOG: An International Journal of Obstetrics & Gynaecology, 122(2), pp.183-189.)

An important feature of a regression line is a numerical measure of the its “slope”. The slope quantifies the steepness (i.e., how steep/slanted) of the line. It is calculated as the change on the Y-axis/vertical axis (usually the outcome) for each unit change on the X-axis/horizontal axis (usually the exposure). If the slope is positive, Y increases as X increases. If the slope is negative, Y decreases as X increases.

A picture containing text

Description automatically generated

Schematic diagram of the slope of a regression line (internet image: <https://www.graphpad.com/guides/prism/latest/curve-fitting/images/embim2.gif>)

Further explanations on regression are given in later lectures.