## Lecture September 05

Case-Control Studies [Fletcher Ch10]

Recall [excerpt from Rothman & Greenland] .. there are two primary types of non-experimental studies in epidemiology.

The first, the **cohort study** (also called *the follow-up study* or *incidence study), is* a direct analogue of the experiment; different exposure groups are compared, but (as in Snow's study) the investigator does not assign the exposure.

The other, the incident case- control study, or simply the **case-control study**, employs an extra step of sampling according to the outcome of individuals in the population. This extra sampling step can make a case-control study much more efficient than a cohort study of the entire population, but it introduces a number of subtleties and avenues for bias that are absent in typical cohort studies.

*** e.g. Hennekens and Buring Fig 2-3
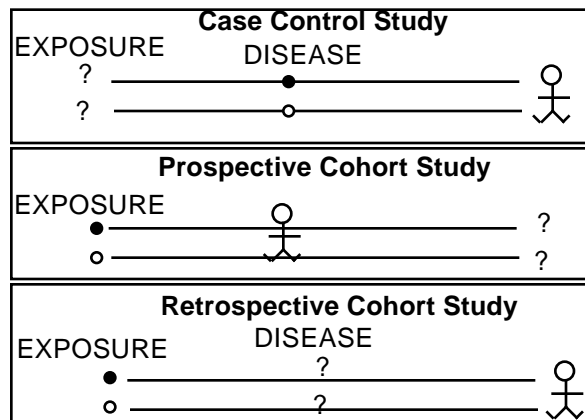
● Present
○ Absent ⟩ basis on which groups are selected at beginning of study

? to be determined   ⚲ investigator at beginning of study

Case-control studies are best understood by defining a source population, which represents a hypothetical study population in which a cohort study might have been conducted. If a cohort study were undertaken, the primary tasks would be to identify the exposed and unexposed denominator experience, measured in person-time units of experience or as the number of people in each study cohort, and then to identify the number of cases occurring in each person- time category or study cohort. In a case-control study, the cases are identified and their exposure status is determined just as in a cohort study, but denominators from which rates could be calculated are not measured. Instead, a control group of study subjects is sampled from the entire source population that gives rise to the cases.

The purpose of the control group is to determine the relative (as opposed to absolute) size of the exposed and unexposed denominators within the source population. From the relative size of the denominators, the relative size of the incidence rates (or incidence proportions, depending on the nature of the data) can be estimated. **Thus, case- control studies yield estimates of relative effect measures.** Because the control group is used to estimate the distribution of exposure in the source population,

In sum, case-control studies of incident cases differ from cohort studies according to how subjects are initially selected. A cohort study identifies and follows a population or populations to observe disease experience; a case- control study involves an additional step of selecting cases and controls from this population. [end of excerpt]

NOTE[JH] The statistical precision of the ratio measure of risk is largely a function of the number of cases. The same amount of person time is needed to generate a given no. of cases in a cohort study as in a case-control study. The latter's efficiency derives from the reduced amount of data-gathering, and the underlined investigator's time-scale -- IF the exposure of past cases and "non-cases" can be accurately established after the fact.

The essential difference can be illustrated using the data from John Snow's investigation

"According to a return which was made to Parliament, the Southwark and Vauxhall Company supplied 40,046 houses from January I to December 31, 1853, and the Lambeth Company supplied 26,107 houses during the same period; "

So, the **denominators** were...

No. of Houses with...

	Water	
	Impure	Pure
	40 046	26 107

286 fatal attacks of cholera took place, in the first four weeks of the epidemic, in houses supplied by the former company, and only 14 in houses supplied by the latter

No. of CASES (**numerators**) in houses with...

	Water	
	Impure	Pure
	286	14

Attack **rates** in houses with...

	Water		
	Impure	Pure	**Ratio**	Difference
	$\dfrac{286}{40046}$	$\dfrac{14}{26\,107}$		
	71.4 / 10K	5.4 / 10K	13.3	66 / 10K

This is the cohort approach -- start with denominators of known sizes and then determine the numerators.

But what if the sizes of the two denominators were not readily available (but the numerators were) ???. it would be a lot of leg work to determine the water source of each of 40046 + 26107 = 66153 houses!

1

No. of CASES (**numerators**) in houses with...

Water

Impure	Pure
286	14

If is a huge amount of work to determine the sizes of the two denominators, how about we take a sample and estimate their estimate their <u>relative</u> sizes ?

Say we survey **100** houses selected at random; we might find that the **sources** were...

No. (± sampling variation) of 100 **sampled Houses** with...

Water

Impure	Pure	
61 (±10)	39 (±10)	100

We can take the 61 and 39 as "quasi-denominators" and make two "quasi-rates"

Quasi-attack **rates** in houses with...

Water

Impure	Pure	**Ratio**	Difference
286	14	13.1	no
61	39	(±)	meaning

Lets say that instead we survey **1000** houses selected at random and that the sources were...

No. (± sampling variation) of 1000 **sampled** Houses with...

Water

Impure	Pure
605 (±32)	395 (±32)

Quasi-attack **rates** in houses with...

Water

Impure	Pure	**Ratio**	Difference
286	14	**13.3**	no
605	395	**(±)**	meaning

Thus the purpose of the 100 (or 1000, or however many are selected, depending on the budget, and the statistical precision required) houses selected at random is to determine the relative (as opposed to absolute) size of the exposed and unexposed denominators within the source population. From the relative size of the denominators, the relative size of the incidence rates (or incidence proportions, depending on the nature of the data) can be estimated.

A good descriptor of these houses selected at random is "the **denominator series**". The cases, already in hand, constitute the "numerator series". [terminology of McGill Prof Miettinen]

To make the calculation of the statistical errors associated with the estimated ratio less complicated, most epidemiologists would exclude the "case houses" from the sampling frame of 66153 houses and would instead sample the "source to be determined" houses from the remainder - i.e. from the "non-case houses". See for example Fletcher et al.'s Figure 10.3, where they write of "non-cases".

Unfortunately, the more common (and older) name for these "non-case" houses is the "control" houses. This creates considerable **confusion** among non-epidemiologists, since we now have 2 meanings for "control" ..

1 in an experiment (e.g. clinical trial), those who do not receive the experimental (new) treatment are sometimes referred to as the "controls" ("comparison group" or --if it is the situation -- "unexposed group" is a more informative label) The same applies in a (non-experimental) cohort study (e.g. what should one call the wives of the male resident physicians when their pregnancy outcomes are compared with those of the female resident physicians?)

Notice that Fletcher et al. themselves use confusing terminology -- in describing the characteristics of a cohort study (Table 10.2 3rd row, 1st column) they say "Controls, the

comparison group (i.e. noncases), not selected -- evolve naturally.

2 in a "study that relies on quasi-denominators", (commonly known as a "case-control" study), the "controls" are the denominator series. Their exposure status (or exposure history) is the focus of the inquiry. Even though it is not entirely accurate, it is less confusing to call them "non-cases" than to call them "controls".

"Being epidemiologically correct"... Most epidemiology textbooks still describe case-control studies as "comparing cases with controls". In fact, as the above example [that views the "controls (or non-cases) as a denominator series] shows, e**ven in a case-control study one compares** (quasi-rates) for the **exposed** with quasi-rates for the **non-exposed** (in the ratio of these quasi-rates, the hidden sampling fraction cancels out in the arithmetic)

This last point about the sampling fraction is very important: the "controls" [i.e., the "non-case" or "the denominator series"] must be selected without regard to their exposure.. see page 1 re "this **cardinal requirement"**

Other simple e.g.'s of denominator issue:

"Pour battre Patrick Roy, mieux vaut lancer bas" (JH course 626)

WOMEN ARE SAFER PILOTS: newspaper article (JH course 626)

Could we use a case-control approach to the Study of Medical students' compliance with simple administrative tasks and success in final examinations?

Most important issues in case-control studies

- appropriate "controls"
    "nested" case-control approach attractive
    i.e. explicit source (e.g.. Medicare database of all Saskatchewan residents: tumor registry; database of all prescription drugs dispensed [universal drug coverage])

- selection bias (e.g. naive md's contributed   only the "<u>exposed</u> cases" in study of 3rd generation OC's and risk of venous thromboembolic disorders)

- exposure and confounder ascertainment