

8. Analytic study designs

The architecture of the various strategies for testing hypotheses through epidemiologic studies, a comparison of their relative strengths and weaknesses, and an in-depth investigation of major designs.

Epidemiologic study designs

In previous topics we investigated issues in defining disease and other health-related outcomes, in quantitating disease occurrence in populations, in relating disease rates to factors of interest, and in exploring and monitoring disease rates and relationships in populations. We have referred to cohort studies, cross-sectional, and case-control studies as the sources of the measures we examined, but the study designs themselves were secondary to our interest. In the present chapter we will define and compare various study designs and their usefulness for investigating relationships between an outcome and an exposure or study factor. We will then examine two designs – intervention trials and case-control studies – in greater depth.

The study designs discussed in this chapter are called analytic because they are generally (not always) employed to test one or more specific hypotheses, typically whether an exposure is a risk factor for a disease or an intervention is effective in preventing or curing disease (or any other occurrence or condition of interest). Of course, data obtained in an analytic study can also be explored in a descriptive mode, and data obtained in a descriptive study can be analyzed to test hypotheses. Thus, the distinction between "descriptive" and "analytic" studies is one of intent, objective, and approach, rather than one of design. Moreover, the usefulness of the distinction is being eroded by a broad consensus (dogma?) in favor of testing hypotheses. Since to characterize a study as "only descriptive" tends to devalue it, investigators understandably try to portray their studies as "analytic" and "hypothesis-driven" in order to make a better impression and to improve their chances for funding and journal space. (Opinions expressed herein are not necessarily those of the sponsor!)

Whether the study is "descriptive" or "analytic", it is important to clearly identify the objectives of the study (preferably identifying the specific parameters to be measured – see Rothman and Greenland) and the rationale (i.e., the case for conducting the research). There are innumerable decisions, judgments, and compromises that must be made during the design, conduct, analysis, and interpretation of a study, and the principal guideposts for making them are the study objectives and rationale. For example, if the objective is to test hypotheses, then the investigator designs and conducts the study so as to maximize the usefulness of the data for testing these hypotheses. Failure to keep the study objectives prominently in one's mind increases the advantage of hindsight over foresight.

Epidemiologic investigations of disease etiology encounter many challenges, especially when they must contend with one or more of the following:

1. Difficulties in defining and measuring the disease;

2. Imprecision in determining the time of onset of the disease;
3. Prolonged intervals between exposure to a causative agent and disease onset (induction period) and between disease onset and detection (latency);
4. Multifactorial disease etiology; and
5. Differential effect of factors of interest on incidence and course of the disease.

[See Mausner and Kramer, Chapter 7, pp 178 *et seq.*]

Even more daunting can be studies of phenomena other than clinical diseases, where less assistance is available from the biomedical armamentarium.

In view of these and other challenges, including the logistical and practical ones of obtaining access to subjects, measuring variables of interest, protecting subjects' rights, and assembling sufficient cases for rare diseases, the basic epidemiologic analytic strategy may be characterized as "by any (ethical) means necessary", along with "try to get the best but if you have to, make do with what's available". For this reason there are innumerable variations in the details of study design. But in terms of the basic architecture - how the principal components of a study are assembled - there are certain basic designs.

Traditional classification of epidemiologic study designs

A logical sequence of study designs encountered in epidemiology is:

1. Case reports
2. Case series
3. Ecologic (also called correlational)
4. Cross-sectional
5. Case-control
6. Follow-up/cohort
7. Intervention trials/controlled trials

The first two of these designs are employed in clinical, rather than epidemiologic, studies, but often are precursors to epidemiologic studies. The next two designs are regarded as primarily descriptive, the last design is primarily analytic, and designs 5 and 6 can be employed in analytic (hypothesis testing) or descriptive modes, depending upon the extent to which the study is oriented towards a pre-existing specific hypothesis. Of course, it may be difficult to obtain resources for a lengthy or expensive study without *a priori* hypotheses, but there are exceptions. Of course, once the data have been collected for whatever purpose, they will often be subject to a search ("search and destroy", as some would have it; "seek and ye shall find" in the view of others) for other associations and insights.

Progression of types of studies

In the classic or ideal scenario, studies of disease etiology unfold from simple, inexpensive, and rapid investigations that identify hypotheses to complex, costly, and lengthy ones to evaluate these hypotheses. General, exploratory studies typically take place before highly focused studies.

New syndrome or outbreak

The stimulus to investigating disease etiology may be prompted by the appearance of a new or previously unrecognized syndrome. In this case the initial efforts will be aimed at characterizing the syndrome, developing a case definition, and searching for characteristics that differentiate people with the disease from persons without the disease. Or, a previously recognized disease may occur in a population group or geographical area where it has not been thought to occur. Such nonroutine situations then prompt a case report in a medical journal, notification of public health officials, or other actions that lead to initial studies – typically case series's and outbreak investigations – to define the nature of the situation and to look for leads to its cause.

As we recounted in an earlier chapter, the history of AIDS epidemiology followed this classic pattern. Recognition of AIDS began with case reports and case series's describing cases of young otherwise healthy men in California and New York City with *Pneumocystis carinii* pneumonia (PCP) and Kaposi's Sarcoma (*MMWR* 1981;30:250-2 and 305-8). Before that time, PCP had been seen only in persons who had been medically immunosuppressed in connection with a transplant operation. Kaposi's Sarcoma had been known as a disease of Africans and elderly men of Mediterranean origin. The initial case series's described common and variable features of the syndrome. For example, all of the patients were men who had sex with men, most had a large number of male sex partners, and many used inhalants, a type of recreational drugs.

The case series's led to an initial AIDS case definition for the purposes of identifying additional cases and inaugurating surveillance. With a case definition in hand, it was also possible to conduct case-control studies in which persons with the disease could be compared with persons without the disease and characteristics associated with the condition identified. Comparisons of AIDS cases to apparently healthy male homosexual controls indicated that the cases had higher numbers of partners, had greater involvement in certain sexual practices (anal intercourse, fisting), and more exposure to drugs used to enhance sexual pleasure. These findings led to analytic studies to test these and other exposure hypotheses.

Case reports and case series's are the clinical route to definition and recognition of disease entities and to the formulation of hypotheses. These studies are not "epidemiologic" in the sense that they have no explicit comparison group or population reference. On the other hand, one can think of an implicit comparison with "common knowledge", "general experience", etc., when the characteristics of cases are striking. An example is history of maternal exposure to diethylstilbesterol (DES) in teenage women with vaginal adenocarcinoma. Other diseases where the clinical route to hypothesis development was prominent are dental caries and fluoride, congenital malformations later linked to maternal rubella infection and retrolental fibroplasia in premature newborns later linked to oxygen exposure.

Sometimes the appearance of a new syndrome is sufficiently alarming that public health authorities are notified and involved at the outset. For example, toxic shock syndrome, with its rapid and malignant clinical course, Legionnaire's disease, where a group of conventioners became severely ill within hours of one another, and rapidly fatal Hanta virus infections among American Indians living in the Southwestern United States in 1994 prompted investigations by public health authorities thereby prompting a much more intensive investigation of microbiologic and environmental factors.

Descriptive studies and surveillance

An alternate stimulus to investigation may come from a surveillance activity or descriptive study. The descriptive study might be a re-analysis of data collected for some other purpose (e.g., from a national population survey or possibly from an analytic study of another hypothesis or even another disease), a mapping study in which disease rates are plotted geographically, or an "ecological" study that uses data on populations rather than on individuals. For example, Warren Winklestein's observation that in the Third National Cancer Survey (US) geographical areas with high rates for cervical cancer tended to have high rates for lung cancer led him to the hypothesis that cigarette smoking might be a risk factor for cervical cancer.

Observations made from population-level data require additional caution in their interpretation, however. For example, colon cancer rates are higher in U.S. counties that use mostly surface water and in countries with high per capita meat consumption. These relationships suggest that something about surface water, e.g., chlorination, and something about meat consumption, e.g., saturated fat intake, might be factors in the development of colon cancer. However, since exposure is not known at the individual level, it is possible that the cases of colon cancer are not themselves people who drink chlorinated water or eat meat. The attempt to infer individual characteristics or relationships from group-level measures is called the "ecologic fallacy". Ecologic, or group-level, studies can nevertheless contribute important information, though, and not only in an exploratory mode.

Once the hypothesis has been advanced, analytic studies are the next epidemiologic recourse. The progression of designs at this point depends on the nature of the disease and exposure - the rarity of the disease, the length of its natural history, the problems in measuring disease and exposure, and other factors. For many diseases, especially rare ones, the usual sequence is to begin with case-control studies (since these are generally the most efficient and logistically practical design) and, unless negative results occur and are accepted, move towards follow-up studies and possibly intervention studies.

Individual-level studies

Although an "epidemiologic transition" appears to be underway, most analytic studies have the person as the unit of data collection and analysis. Thus, the four classic analytic study designs are generally thought of in relation to individual-level studies, though as we shall see they can also be employed for studies where the group is the unit of analysis. These four primary designs are:

Cross-sectional

A cross-sectional study is one in which subjects are sampled without respect to disease status and are studied at a particular point in time, as in a random-sample health survey. The term "cross-sectional study" (or "prevalence study") usually refers to studies at the individual level, even though ecologic studies are typically (though not necessarily) cross-sectional, also. The target population is generally one whose identity is of some wider interest (e.g., a political or geographical entity, a profession or workforce, or a major organization (union, HMO, student body), but may not necessarily be so.

In a cross-sectional study, the current or historical status of individuals is assessed and may be examined in relation to some current or past exposure. These studies are obviously most useful for conditions that are not rapidly fatal, not terribly rare, and/or not routinely brought to medical attention (e.g., elevated blood pressure, elevated blood cholesterol, many psychiatric disorders, diet, subclinical infection, and serologic markers of previous infections).

Since participants for a cross-sectional study are generally chosen without previous knowledge of their disease or exposure status, such studies can be used to estimate prevalence of both diseases and exposures and therefore to compute prevalence ratios and prevalence odds ratios.

Among the more widely known cross-sectional studies are the periodic national household (interview) surveys by the U.S. National Center for Health Statistics (NCHS), the annual (telephone) Behavioral Risk Factor Survey conducted by the U.S. Centers for Disease Control and Prevention (CDC), and HIV seroprevalence studies. Sometimes the process of recruiting subjects to a follow-up study (e.g., the Lipids Research Clinics Coronary Primary Prevention Trial prevalence study) serves as a cross-sectional study. The cross-sectional NCHS NHANES (National Health and Nutrition Examination Survey) study became a follow-up study when respondents were re-examined ten years later, creating the NHANES Follow-up Study.

Strengths

- Can study entire populations or a representative sample.
- Provide estimates of prevalence of all factors measured.
- Greater generalizability.

Weaknesses

- Susceptible to selection bias (e.g. selective survival)
- Susceptible to misclassification (e.g. recall)
- Information on all factors is collected simultaneously, so it can be difficult to establish a putative "cause" antedated the "effect".
- Not good for rare diseases or rare exposures

Case-control (case-referent, etc.) studies

A case-control study is one in which persons with a condition ("cases") are identified, suitable comparison subjects ("controls") are identified, and the two groups are compared with respect to prior exposure. Thus, subjects are sampled by disease status. Case-control studies are used in infectious disease epidemiology, but they have become the primary strategy in chronic disease epidemiology. The investigation and refinement of the case-control design, a process which began in about the middle of the 20th century (see classic articles by Cornfield, 1951 and Mantel and Haenszel in 1959) constitutes a significant innovation in population-based research. (Note: The analogy that presumably led case-control theorists to adopt the term "control" from experimental designs is accurate only in a general sense, i.e., in both cases the control group serves as a point of reference of comparison for the group of primary concern. However, because of the fundamentally different architecture of experimental and case-control designs, the analogy ends there and has probably been a source of confusion in earlier writings about the case-control design. See the end of the section on selection bias in the next chapter.)

Because subjects are identified after the disease has developed, and inquiry then investigates prior exposure, the case-control study is sometimes referred to as a "retrospective" or "backwards" design. The "backwards" design poses greater demands in terms of methodological and analytic sophistication. However, by ensuring a greater balance between the numbers of cases and noncases, the case-control design generally offers much greater statistical efficiency than other designs, giving it a crucial advantage for studying rare diseases.

Case-control studies can use prevalent cases (i.e., existing at the time the study begins) or incident cases (i.e., newly diagnosed during the period of the study). In the former instance, the distinction between a case-control study and a cross-sectional study can become very blurred. In addition, data collected through other kinds of studies can be analyzed as if data had come from a case-control study, thereby providing another source of confusion.

Because case-control studies select participants on the basis of whether or not they have the disease, the case-control design does not provide an estimate of incidence or prevalence of the disease, unless data about the population size are available. But as long as the participants are chosen without regard to their exposures, the study can estimate the prevalence of one or more exposures. With these prevalences, in turn, we can estimate an exposure odds ratio which we then use to estimate the IDR or CIR in the base population.

Strengths

- Good for rare diseases
- Efficient in resources and time

Weaknesses

- Susceptible to selection bias (e.g., cases or controls may not be appropriately "representative")
- Susceptible to misclassification bias (e.g. selective recall)
- May be difficult to establish that "cause" preceded "effect".

Follow-up studies

Along with case-control studies, follow-up studies constitute the other basic observational strategy for testing hypotheses. In a follow-up study, people without the disease are followed up to see who develops it, and disease incidence in persons with a characteristic is compared with incidence in persons without the characteristic. If the population followed is a defined group of people (a "cohort"), then the study is referred to as a cohort study. Alternatively, the population under study may be dynamic (e.g., the population of a geographical region).

Follow-up studies may be done "retrospectively", where the population at risk can be defined at some time in the past and traced forward in time, or "prospectively", where the population is identified or assembled by the investigator and then followed forward in time.

Since the study population for a follow-up study is selected from among people who are free of the disease, this study design can estimate incidence based on new cases that develop during the follow-up period. Because the investigator can estimate incidence separately for exposed and unexposed participants, the IDR and/or CIR can be directly obtained from the incidence estimates. In some cases, the study population is gathered on the basis of an initial cross-sectional study (e.g., the Framingham and Evans County cohorts). In such cases, exposure prevalences in the base population can also be directly estimated, though this ability comes from the cross-sectional component, not from the follow-up component.

Strengths

- Better for rare exposures
- Less confusion over relative timing of exposure and disease than with other observational designs.

Weaknesses

- Costly and time consuming if disease is rare and/or slow to develop.
- Loss to follow-up (attrition) may lead to selection bias.
- Relatively statistically inefficient unless disease is common.

Intervention trials (controlled trials)

An intervention trial is a follow-up study in which the primary exposure under study is applied by the investigator. These are the only experimental form of epidemiologic studies, though they are also observational in that subjects remain in their ordinary habitats. In an intervention trial, the investigator decides which subjects are to be "exposed" and which are not (in contrast to naturalistic studies in which the subjects "choose" their exposure group by "deciding" whether to smoke, drink, exercise, work in a hazardous environment, be exposed to toxic wastes, breathe polluted air, develop elevated blood pressure, develop diabetes, etc.).

The term "clinical trial" emphasizes the controlled aspect of the intervention, at the expense of the generalizability of the results; the term "community trial" emphasizes that the trial is carried out in a

realistic setting and results may therefore be more generalizable (at the expense of having control over what subjects actually do). A community trial can involve an individual-level intervention (e.g., breast cancer screening), a community-level intervention (e.g., gun control), or interventions with elements of both levels (e.g., mass media promotion of physical exercise).

In the United States, the National Heart Lung and Blood Institute (NHLBI) sponsored and led several major (thousands of subjects, multiple expensive follow-up examinations, many millions of dollars) individual-level randomized intervention trials to confirm the value of modifying coronary heart disease and cardiovascular disease risk factors: the Hypertension Detection and Follow-up Program (HDFP), Multiple Risk Factor Intervention Trial (MRFIT), and the Lipids Research Clinics Coronary Primary Prevention Trial (LRC CPPT). More recently, the National Cancer Institute (NCI) began large-scale trials to assess effectiveness of screening techniques for cancers at a number of sites (colon, prostate). Probably the largest individual-level randomized trial in the U.S. is the Women's Health Initiative (WHI) which is funded through the National Institutes of Health (NIH, of which both NHLBI and NCI are subdivisions). Large trials of this type have also been conducted in Australia, Canada, Europe, and probably elsewhere that I am not yet aware of.

Intermediate between a formal intervention trial and a follow-up study are follow-up studies in which the intervention is applied by an outside agency (e.g., a health care provider or organization) but is not being manipulated in response to an experimental design.

Strengths

- Most like an experiment
- Provides strongest evidence for causality in relation to temporality and control for unknown "confounders"
- Fulfills the basic assumption of statistical hypothesis tests

Weaknesses

- Expensive, time consuming, sometimes ethically questionable.
- Subjects are often a highly selected group (selected for willingness to comply with treatment regimen, level of health, etc.) and may not be representative of all people who might be put on the treatment (i.e., generalizability may suffer).

Group-level (ecologic) studies or measures

Group-level studies (also called ecologic studies, correlational studies, or aggregate studies) obtain data at the level of a group, community, or political entity (county, state, country), often by making use of routinely collected data. When they use data that are already available and usually already summarized as well, these studies can be carried out much more quickly and at much less expense than individual-level studies. Group-level studies may also be the only way to study the effects of group-level constructs, for example, laws (e.g., impact of a seatbelt law), services (availability of a suicide prevention hotline), or community functioning. Multi-level studies can include both individual-level (e.g., disease, individual exposure) and group-level (e.g., median family income) variables at the same time. The popularity of multi-level studies is growing rapidly, due to the return

of interest community-level influences and the increasing availability of statistical algorithms and software to analyze multilevel data.

Each of the four classical study designs discussed above (cross-sectional, case-control, follow-up, intervention) can also be carried out with group-level variables. Thus, a set of counties, states, or countries can be analyzed in a cross-sectional manner to look at the variation in a health variable (e.g., mean blood pressure, hospitalizations for asthma, homicide rates, imprisonment rates) and its relationship to country characteristics (e.g., salt intake, air pollution, handgun laws or possession, drug policies). Many group-level studies are of this type. (Studies of homicide rates, new hospitalizations, and other phenomena that represent events, rather than conditions, should perhaps be regarded as follow-up studies, rather than cross-sectional. When only a single year's data are being analyzed or when data for several years are combined into an annual average, the traditional perspective has been cross-sectional.)

Similarly, an investigator can assemble a set of groups (e.g., animal herds, states) with high rates of some health outcome and compare their characteristics with those of states with low rates, as in a case-control study, or can monitor aggregate populations as in a follow-up study to see if differences in baseline variables (e.g., restrictions on cigarette advertising, higher cigarette taxes) are reflected in the development of outcomes (smoking initiation by adolescents). Finally, a group-level intervention trial can be conducted in which schools, worksites, neighborhoods, or political subdivisions are assigned to receive an interventions (school health clinics, curricula, media messages, or lay health advisor programs) and outcomes are monitored over time. Among the more widely-known community intervention trials are the National Cancer Institute COMMIT trial (for smoking cessation and prevention), the Stanford Three Community and Five-City Studies (cardiovascular disease), the North Karelia Study (cardiovascular disease), and recent HIV prevention trials using mass treatment for curable sexually transmitted diseases.

One situation where ecologic data are particularly useful is that where a powerful relationship that has been established at the individual level is assessed at the ecological level in order to confirm its public health impact. If a risk factor is a major cause of a condition (in terms of population attributable fraction as well as strength of association), then a lower presence of that factor in a population should presumably be linked to a lower rate of the associated outcome. Examples of studies where this approach has been taken include studies of oral contraceptive sales and CVD in women (Valerie Beral), incidence of endometrial cancer and prescription data for replacement estrogens, and motor vehicular fatalities and occupant restraint legislation or enforcement.

Ecologic measures as surrogates for individual measures

Recent articles have clarified discussions about ecologic studies by noting that there are in fact two basically different types of group-level studies, or, equivalently, two different ways in which a study can be "ecologic" (Charles Poole, *Ecologic analysis as outlook and method*, 1994). In the first type, a study may be "ecologic" in that the exposure status (fat intake for individuals) is estimated from the group average (per capita fat intake). In this case the group-level variable serves as a proxy for the values for individuals. The group-level average is an inferior measure of the values of individuals, but it is often much easier and economical to obtain. In addition to the loss of precision that results from using the group average as the data for individuals, there is also the danger of the

"ecologic fallacy", the erroneous inference that specific individuals in a group share the characteristics of the group.

Of course, most individuals in a group must share the characteristics of the groups which they comprise. But groups are heterogenous, and a subgroup of individuals can easily differ greatly from the group mean. For example, data showing that areas with higher average residential radon levels had higher lung cancer rates than areas with lower levels do not logically imply that the higher lung cancer rate are due to the higher radon levels. Such an inference is based on the ecologic fallacy, because it is possible that the excess lung cancers occurred to people in houses with low radon levels. In that case the group-level average would be an invalid surrogate for individual-level measurements. But even though it is not valid to infer from these data that radon exposure contributes to the elevated lung cancer rates, that may nevertheless be a correct characterization of the phenomenon. Other data are needed to draw the inference; in the meantime, these ecologic data provide the rationale for more in-depth study.

Ecologic measures as the relevant constructs

A second way in which a study can be "ecologic" is if the population, rather than the individual, is the real unit of study. In this case, a group-level factor is itself the exposure (e.g., an anti-smoking ordinance, crime rate, population density) or, occasionally, the disease (e.g., homicide rate). Although epidemiology has a long tradition of using population-level data for descriptive purposes, the use of group-level data for hypothesis testing has been out of favor because of the problem of the ecologic fallacy (even though it applies primarily to the other type of ecologic study), major limitations in the ability to control for the effects of known determinants of the outcome under study, and the ascendancy of the biomedical paradigm in conjunction with the enormous expansion in capabilities for biochemical measurement and analysis.

How one regards ecologic studies depends to a certain extent on which type of studies are being considered - studies in which group-level variables are measured as economic and convenient, but inferior, measures of diseases and exposures at the individual level or studies in which the phenomena under study operate at the level of the group, rather than (or as well as) the individual. A major modifying influence, though, is one's perspective on epidemiology and public health (see chapter "The role of epidemiology in public health"). In Charlie Poole's (*AJPH*, May 1994) formulation, epidemiologists who regard the health of a community as more than the summation of the health of its individual members, regard ecologic studies (of the second variety) as critical to conduct. In contrast, epidemiologists who regard the health of a community as the summation of the health of its members regard individual-level studies as the superior form of investigation.

Although the latter view remains the dominant one in the U.S. epidemiology profession and government funding for epidemiologic research, the former has been gaining renewed attention, as evidenced by the series of articles in the May 1994 *American Journal of Public Health* from which this section draws heavily. For Susser (who at that time was editor of *AJPH*, though not for his articles), the prime justification for the ecological approach in epidemiology is the study of health in an environmental context: pairings, families, peer groups, schools, communities, cultures – contexts that alter outcomes in ways not explicable by studies that focus solely on individuals (The logic in ecological: I. The logic of analysis. *AJPH* 1994). And where group-level constructs are involved, the

ecological approach may be the appropriate level of study (Schwartz *AJPH* 1994; Susser *AJPH* 1994 [both his articles]).

Multi-level studies

Multi-level studies provide an area of agreement in this debate, since they potentially combine the advantages of both individual- and group-level studies. By using sophisticated methods of analysis – which are only now starting to become readily available thanks to the computer revolution and the development of statistical software – investigators can create mathematical models that include both group-level and individual-level variables. In principle, then, the investigator can take advantage of the ability to control for individual variability and the measurement power and precision offered by biochemical technology while at the same time addressing social, economic, and institutional influences at the community-level.

But such advantages come at a cost. Studying the effects of a group-level variable requires data for a large enough number of groups to enable comparison among them. Routinely collected data (e.g., census data) make such studies economical and relatively easy to conduct. A multi-level study, however, requires individual-level data as well, which typically means primary data collection with its attendant costs, challenges, and time. Moreover, the individual-level data must now be obtained from individuals in a larger number of groups (e.g., worksites, counties) than might be necessary if the objective of the study focused on individual-level variables.

Types of group-level variables

Group-level variables do not all possess the same degree of "groupness". One variety of group-level variable is are summaries of individual characteristics, such as per capita income. Such a variable has been termed contextual (Mervyn Susser, *The logic in ecological: I. The logic of analysis. AJPH* 1994) or aggregate (Hal Morgenstern, chapter 23 in Rothman and Greenland). Such variables illustrate the distinction between individual-level and group-level perspectives, since the aggregate variable measures a different construct from its name-sake at the individual level (Schwartz *AJPH* 1994). Thus, per capita income may be used as a surrogate measure of individual or family socioeconomic status, in which case it is inferior to the individual-level measure, or may instead directly measure income at the community-level, in which case it is a group-level measure with implications for availability of goods, services, facilities, and opportunities of all kinds education, commercial vitality, neighborhood safety, and many other aspects of the social and institutional, and physical environment.

Variables that are not summary measures of individual-level variables include factors like climate, air pollution, disasters, and laws. Susser uses the term integral variable for a variable that does not have a corresponding individual-level value. Integral variables, according to Susser, cannot be analyzed at the individual level.

Morgenstern differentiates between environmental measures and global measures. Environmental measures are "physical characteristics of the place in which members of each group live or work (e.g., air-pollution level and hours of sunlight)" and which have individual-level analogs whose value

can vary substantially among individuals. In contrast, global measures have "no distinct analogue at the individual level . . . (e.g., population density, level of social disorganization, the existence of a specific law, or type of health-care)" (p460).

One can imagine, though, that global measures may also affect individuals differently. Thus population density affects people in different ways depending upon their occupation, preferred activities, transportation requirements, needs for services, and economic resources. Social disorganization affects people more or less depending upon their age, personal social networks, occupational affiliations, need for social services, and, of course, economic resources. Anatole France's aphorism that the law forbids both the poor and the rich alike from sleeping under a bridge or stealing a loaf of bread reminds us that the law does not affect all individuals in the same way. The individual-level effects of the type of health care system depends upon the individual's need for health services, mobility, and, of course, economic resources. Even climate presumably has weaker effects on people with good climate control in their home, workplace, and automobile and who can take extended vacations.

Dependent happenings

An important category of contextual variable is "dependent happenings", where a phenomenon propagates from one person to others. Dependent happenings arise most obviously in the case of contagious diseases, where the prevalence is both a summary of individual infection status but also greatly affects the risk of infection for exposed, nonimmune persons. As an example of the inability of individual-level analysis to analyze a situation with dependent happenings, Koopman and Longini (*AJPH* May 1994;84:836-842) present a study of dengue fever in Mexican villages. The study, carried out following a multi-year epidemic, examined the association between history of infection (measured by antibody test) and presence of *Aedes aegypti* larvae in a household. The odds ratio for an analysis at the individual level was 1.1, i.e., presence of larvae was not related to a positive antibody test. By contrast, the ecological (village-level) analysis yielded an OR of 12.7.

The authors' explanation for this difference is that transmission (i.e., dependent happenings) decreases individual-level effects and increases ecological effects. With a sufficient number of infected persons in a village, the mosquitoes carry the infection to others in that village, even those whose household has not been a breeding ground for mosquitoes. In a village with few infected persons, the mosquitoes are less likely to acquire the virus so households with larvae are not in fact at elevated risk. In this scenario, higher infection prevalence in a village contributes to the ecological relationship directly (because infection prevalence is the outcome variable) and indirectly (in that mosquitoes in high prevalence villages are more likely to get infected).

Other phenomena and situations can also obscure effects of risk factors for transmission in individual-level studies (Koopman and Longini, citing Koopman *et al.* 1991). In fact, when a risk factor affects transmission, neither individual-level analysis nor ecological analysis works. Although infectious diseases have received the greatest attention in such work, psychosocial and behavioral phenomena (e.g., drug use including smoking and alcohol, racism) probably also constitute dependent happenings in some regards.

What measures can be estimated from basic epidemiologic study designs?

The beauty of a follow-up study is that the investigator gets to watch what is happening and to summarize the experience by calculating simple measures like the proportion of exposed subjects who develop the disease ("the incidence of the disease in the exposed") or the rate at which the disease develops in the exposed. This is often not the situation in a case-control study, in which the investigator typically assembles cases without identifying the entire exposed and unexposed populations from which the cases arise.

It is said that a follow-up study "samples by exposure status" and a case-control study "samples by disease status". This is certainly true for a case-control study, but not necessarily so for a follow-up study, which can sample without regard to exposure status. A cross-sectional study can sample by either disease or exposure or neither (i.e., a true "cross-section"). When a cross-sectional study samples by existing disease, it is essentially the same as a case-control study with prevalent cases. However, many of these concepts remain the subject of debate (if interested, see references in the first section of the bibliography).

Multiaxial classification of study designs

There have been various attempts to classify study designs in a more analytic fashion than the conventional taxonomy presented in this chapter. One approach, presented in Kleinbaum, Kupper, and Morgenstern's textbook *Epidemiologic research: principles and quantitative methods*, analyzes major designs in respect to "***directionality***" (cohort studies are said to involve "forward directionality", case-control studies to involve "backward directionality", and cross-sectional studies neither), "***timing***" (the chronological relationship between the most recent data gathering and the occurrence of the study factor and disease – if both study factor and disease were established and measured before the study began, then the study was completely "retrospective"; if both study factor and disease have not yet occurred when the study begins, then the study is completely "prospective" so that measurements can be tailored to the study requirements; studies with exposure data collected both before and disease onset studied after the start of the study were "ambispective"), "***type of population***" (cross-sectional or longitudinal, fixed cohort or dynamic population), and "***unit of observation***" (individual-level data, group-level data).

Various other conceptualizations are also in use. For example, sometimes case-control studies are said to involve "sampling on disease", because cases and controls are sampled separately (as in stratified random sampling). From this perspective, cohort studies are said to involve "sampling on exposure" – exposed and unexposed persons are sampled separately. However, though separate sampling may be necessary in order to obtain a large enough number of participants with a rare exposure, if the exposure is not rare then participants can be selected without regard to exposure status.

Participants for a cross-sectional study can be selected without regard to exposure or disease status, separately by exposure status, or separately by disease status. In the last case, a cross-sectional study is equivalent to a case-control study using prevalent cases. A basic point but one worth noting is that a study cannot estimate a dimension that has been set by its design. That is, if participants are selected separately according to their exposure status, then the proportion who are exposed cannot

be estimated from that study since that proportion is determined by the study design (and its success in recruitment), rather than from the sampling process. If participants are selected according to disease status, then exposure proportions (and odds) can be estimated but not disease prevalence (or odds). That is the reason that one cannot directly estimate risk in a case-control study. (Rothman and Greenland use the term "pseudo-risks" to refer to the proportion of cases among exposed and unexposed case-control study participants.)

Design attributes

As can be seen in the bibliography for this chapter, classification of study designs has been the subject of vigorous debate. Nevertheless, there are various important design attributes that should be noted for any given study. These attributes are:

Subject selection

Under this heading come the various considerations used in selecting participants for the study (e.g., restriction to certain age groups, enforced comparability between groups being compared (matching), natural comparability [twins, siblings], random sampling).

Method of data collection

Data can either be primary data, collected for the purposes of the study at hand or secondary data, collected for purposes other than the study at hand, such as from medical records, death certificates, billing records, or other administrative files. Data may have been collected in the distant past.

Unit of observation

As noted above, data can be collected at the individual level or only at the group level.

Evaluation of a study design

The primary dimensions for evaluating the design of a particular study are:

Quality of information: How accurate, relevant, and timely for the purposes of the study are the data?

Cost-effectiveness: How much information was obtained for how much expenditure of time, effort, resources, discomfort, etc.?

[For more on the above, see Kleinbaum, Kupper, and Morgenstern, *Epidemiologic research: principles and quantitative methods*, ch 4-5.]

The following layout may be useful for reflection or discussion, but cannot be completed unambiguously since in many cases the relative strengths of different designs depend upon the particular study question and circumstances.

Strengths and weaknesses of the classic study designs

	Cohort		Case-control	
	(prospective)	(historical)	(incident)	(prevalent)
Ability to estimate risk				
Ascertainment of cases				
(access to care, diagnostic criteria, selective survival)				
Measurement of exposure				
Reliance on historical data				
Selective recall				
Disease may affect characteristic				
Control of all relevant variables				
Study affects subject behavior				
Temporality established				
Feasibility and logistics				
Rare exposures				
Rare diseases				
Statistical power and efficiency				
Attrition				
Time and effort				
Ethical concerns				
Cost				

Individual-level interpretations of measures of association

The individual-level follow-up study, cross-sectional study, and case-control study are fundamental designs in epidemiologic research. Data collected using any of these designs allow one to estimate an individual-level measure of association or effect, i.e., a measure of the strength or magnitude of the quantitative relationship of a study factor (i.e., exposure of interest) with a disease. We learned about these measures in a previous chapter. We revisit them here to reinforce the relationship between which measures can be estimated with which study designs.

One way of conceptualizing study designs is to regard the objective of an etiologic individual-level study as the estimation of a measure of effect relating an exposure to a disease outcome, specifically a risk ratio (CIR) or rate ratio (IDR). The preference for these measures is that, as Greenland (1987) demonstrates, they are interpretable at the level of the individual's risk or hazard function so that under certain assumptions an RR of two means that an exposed individual's risk or hazard is twice that of an unexposed individual. (Although the odds ratio does not possess an interpretation in terms of an individual's odds, it is useful through its ability to estimate a risk ratio or rate ratio. Similarly, the prevalence odds ratio is of interest primarily because under certain assumptions it estimates the incidence density ratio (rate ratio) [Greenland, 1987]).

Risk ratio

Consider the example of a pregnant woman who drinks three or more alcoholic drinks per day during pregnancy. Suppose that that drinking that amount of alcohol is associated with a 20% chance of bearing a malformed baby. If that chance is 2% for a pregnant woman who does not drink, the ratio of fetal malformations in relation to drinking three drinks/day is 10 (20%/2%). A risk ratio of 10 indicates a very strong association and therefore one that is more likely to be causal. Also, the relative risk conveys a clear, intuitive meaning about the degree by which the exposure increases risk.

We can also interpret this risk ratio at the individual level: the risk for an individual woman who drinks 3+ alcohol drinks/day during pregnancy was 10-times (or 900% greater than) that for a woman who does not drink. Such an interpretation, of course, involves a number of assumptions, i.e., that apart from the effect of drinking, the women in the exposed group have the same risk as women in the unexposed group and that the individual woman to whom the group-level association is being imputed has risk-related characteristics close to the group average. But mathematically there is no problem. [Aside: Birth outcomes such as fetal malformations are generally regarded as prevalences among babies born, since the denominator for births is generally unknowable; for simplicity the above example assumes that all pregnancies result in a live birth.]

Rate ratio

Often we estimate disease rates, rather than risks, in which case the measure of effect of interest is a rate ratio. For example, in a study of breast cancer in relation to early use of oral contraceptives, we may have anywhere from 10 to 20 years of follow-up on subjects. To accommodate these differing lengths of follow-up, we can calculate the rate of breast cancer cases per woman-year, rather than per woman. In that case a two-fold elevation would mean that the rate at which breast cancer cases are observed in women with early use of oral contraceptives was twice that in women without early use of oral contraceptives. Again, the rate ratio has an interpretation at the individual level (Greenland, 1987) and can be mathematically converted into an estimate of relative risk over a given time interval. It can also be interpreted in terms of the expected time until the event occurs in the average woman.

Incidence odds ratio

The incidence odds ratio is the ratio of odds of disease in exposed persons to the odds of disease in unexposed persons. Odds are ratios of risks. If the risk is r , the odds are $r/(1-r)$. When the risk is small, risk and odds are nearly equal, and the odds ratio approximates the rate ratio and risk ratio.

Since the odds ratio can be estimated in a case-control study even where no other measure of relative risk is directly available, the odds ratio is of great practical importance for epidemiologists. The prevalence odds ratio (from a cross-sectional study) also approximates the rate ratio when the duration of the condition is unrelated to exposure status. The prevalence ratio can also be the measure of primary interest, when duration is itself the outcome, such as in the treatment of depressive disorder. However, mathematically (see Greenland) there is no direct individual-level interpretation for the odds ratio (whereas the incidence proportion is the sum of the risks across all

individuals, this relationship does not hold for the incidence odds and individual odds). For this reason, Greenland argues, the CIR and IDR are preferred.

Preferred measures of association

So, our primary interest for etiologic purposes is generally the risk ratio (CIR) or rate ratio (IDR). Where we cannot estimate either of those directly, then we usually try to design the study so that we can estimate the odds ratio and use it to estimate the rate ratio or risk ratio. We may also want to estimate a measure of impact, to quantify the importance of the relationship we are studying should it turn out to be causal. In the table below are listed the kinds of measures of association and impact that can be derived from the basic epidemiologic study designs:

Measures of association for basic epidemiologic study designs

Type of study design	Measure of association	Measure of impact
Follow-up, person denominator	Risk ratio	Absolute Relative
Follow-up, person-time denominator	Rate ratio	Absolute Relative
Case-control	Odds ratio	Relative
Cross-sectional	Prevalence odds ratio or prevalence ratio	Relative

Formulas for and examples of computation

Construct (2x2, four-fold) table:

Disease	Exposure		Total	
	Yes	No		
Yes	a	b	m ₁	(a + b)
No	c	d	m ₂	(c + d)
Total	n ₁	n ₀	n	
	(a + c)	(b + d)		

Example: The following are hypothetical data involving subjects who have been determined to be either hypertensive (diastolic blood pressure >90 mmHg) or normotensive (diastolic blood pressure ≤90 mmHg) and were classified into one of two categories of dietary salt intake, high or low.

Dietary Salt Intake and hypertension(Hypothetical)

Hypertension	Dietary salt intake		
	High	Low	Total
Yes	135	160	295
No	180	420	600
Total	315	580	895

If these data came from a follow-up study, then the risk of disease in exposed subjects would be a / n_1 , the risk in unexposed subjects would be b / n_0 , and the risk ratio or relative risk would be:

$$RR = \frac{a / n_1}{b / n_0} = \frac{135 / 315}{160 / 580} = 1.55$$

If these data came from a cross-sectional study, the calculations would be identical except that the data would yield measures of prevalence and a prevalence ratio instead of risk and risk ratio. However, the prevalence odds ratio (see below) would generally be preferred as a measure of association, since under the assumption of no difference in duration of hypertension between high- and low-salt people, the prevalence odds ratio estimates the incidence density ratio in the population.

If these data came from a case-control study, the above calculations would not be meaningful. Since a case-control study samples subjects on the basis of their disease status, proportion of exposed who are cases does not estimate anything. Rather, we need to compute the odds of cases and controls who are exposed! Thanks to the odds ratio, we can estimate the rate ratio in the population from which the cases arose:

Odds of Exposure in Cases (D):

$$\text{Odds} = \frac{\text{Proportion of cases exposed}}{\text{Proportion of cases not exposed}} = \frac{a / (a + b)}{b / (a + b)} = \frac{a}{b}$$

Odds of Exposure in Controls (D):

$$\text{Odds} = \frac{\text{Proportion of controls exposed}}{\text{Proportion of controls not exposed}} = \frac{c / (c + d)}{d / (c + d)} = \frac{c}{d}$$

Odds Ratio (OR)

$$\text{OR}_e = \text{Exposure odds ratio} = \frac{\text{Exposure odds in cases}}{\text{Exposure in noncases}} = \frac{a / b}{c / d} = \frac{ad}{bc}$$

$$\text{OR}_e = \frac{ad}{bc} = \frac{135 \times 420}{160 \times 180} = 1.97$$

Intervention Trials

(An earlier version of this section was written by Joellen Schildkraut, Ph.D.)

In an experiment, a set of observations are conducted under controlled circumstances. In contrast to nonexperimental, observational epidemiologic studies, experimental studies permit the scientist to manipulate conditions to ascertain what effect such manipulations have on the outcome. The objective of an experiment is the creation of duplicate sets of circumstances in which only one factor that affects the outcome varies. An example is laboratory animal studies such as those which evaluate potential carcinogens.

In such studies, the investigator has a great deal of control over the experimental units, their environment, measurements taken, and exposure to the study factors. Even genetic factors can be controlled by using inbred strains of mice. Experiments provide a means to disentangle complex problems in stepwise fashion, to reduce macro-level phenomena into collections of low-level mechanisms. This reductionist approach, made possible by laboratory experimentation, has made possible the remarkable advances in knowledge and technology of the past few centuries. The rub is that not all phenomena are amenable to dissection in this way. Laboratory experimentation on humans is greatly constrained, and extrapolation from animals to humans often problematic. Also, many conditions of interest cannot be manipulated and it is generally impossible to recreate real-life situations in the laboratory.

In epidemiology, intervention trials are the closest analog of a laboratory experiment. What distinguishes intervention trials from other types of epidemiologic studies is the manipulation of the study factor. This manipulation may be governed by random assignment, creating a true experiment, or if not, a quasi-experiment. Randomization offers the greatest opportunity to create groups that are equivalent in all regards, with the corresponding opportunity to isolate the effect of the intervention. The potential for achieving such isolation in a study with nonrandom assignment depends on the ability to adjust for differences in the analysis. Even with good data on all relevant factors adjustment may not be possible. For example, no analytic technique could correct a study where all patients with a better prognosis were assigned a new drug instead of an old drug.

Intervention trials can include testing therapeutic or preventative hypotheses, the estimation of long term health effects, and identification of persons at high risk. Types of interventions include:

- Prophylactic - focus on prevention (e.g. vaccines, cholesterol lowering)
- Diagnostic - focus in evaluation of new diagnostic procedure (e.g. comparison of a less invasive diagnostic procedure to a gold standard, etc.)
- Therapeutic - focus on treatment (e.g. drug testing, evaluation of new surgical technique, etc.)

A randomized clinical trial (RCT) is defined as a prospective study that estimates the effect of an intervention by comparing participant outcomes between randomly assigned treatment and control groups. The major RCTs are multi-center studies in two or more hospitals with a common protocol. The strengths of multi-center studies are more representative patient populations, larger sample size, and shorter study period (or duration of patient intake). Finally multi-center studies enable research on rare diseases.

Drug trials go through several levels of study:

Phase I - early study to determine dose level that is not too toxic (animal studies)

Phase II - efficacy trial to estimate the effectiveness of an agent with specified precision.

Phase III - comparative trial to test whether the new agent is better than the standard or control agent.

Phase IV - for the detection of rare side effects by way of epidemiologic studies or prospective monitoring

Steps of a Clinical Trial

There are three phases in a clinical trial: 1) planning, 2) the trial (data collection), and 3) concluding phase:

1. Planning phase

Study Design

Clinical trials can be randomized controlled studies or nonrandomized studies (quasi-experiments). If the latter they can have concurrent controls (a group or groups that are regarded as similar to the experimental group and whose experience is observed during the same period of time as that of the experimental group), historical controls (a group regarded as similar and for which data are already available), or sometimes no controls.

Randomization is a method for allocation of subjects to intervention and control groups where each subject is equally likely to be assigned to one or the other. Various randomization procedures have been proposed for use in clinical trials. The most frequently used techniques are:

Simple randomization - assignment of people to treatment groups is random, not concerned with other variables

Balanced Block randomization - ensures balance in the proportion of patients assigned to each treatment within each group or blocks of patients entered. (e.g. hospitals in a multicenter study).

Stratified randomization - is used when there are specific factors known to have a significant effect in the outcome of the trial. Separate balanced block randomization schemes are established within each level of the stratified variable or variables.

In a multicenter study, randomization can be stratified by institution because of institutional differences in the patient population, in the overall level of patient care, or in the treatment effect of the institution.

Blindness or masking

- Non-blinded - common in community trials
- Blinded - the observer is aware but the subject is not aware of treatment assignment
- Double blinded - Neither the observer or the subject is aware of treatment assignment
- Triple blinded - The observer, subject, and data analyst are not aware of treatment assignment

Concurrent and non-randomized controls can result in systematic assignment bias and uninterpretable results. Historical controls may not be comparable in terms of patient selection, external environment (even if it is the same hospital), improved diagnostic tests, and unknown factors, but the cost is cheaper and the length of the time to complete the trial is shortened. Evidence for bias in treatment assignment of controlled clinical trials was illustrated in a study by Chalmers et al. (*N Engl J Med* 1983; 309:1358-61):

Type of Study	No. of studies	>=1 significant prognostic variable	Significant Difference in fatality rate
Blinded Randomized	57	14.0 %	8.8%
Unblinded randomized	45	26.7 %	24.4%
Non-randomized	43	58.1 %	58.7%

Sample size estimates are vital to planning the study. The estimated difference in the response variable (outcome of interest), significance level, and noncompliance rate must be factored into the calculation of sample size.

2. Trial phase (data collection)

Screening can be applied to those already admitted to the hospital or those who can be contacted from outpatient services. Patients should be those who are likely to benefit from the intervention and those who are likely to comply with the intervention schedule.

Treatment allocation can be 1) fixed in the beginning, optimally in one to one ratio, 2) it can be adaptive allocation where results of an ongoing trial influences allocation so that the proportion of patients with beneficial treatment is maximized or 3) crossover design which helps to eliminate the variation between patients.

Study monitoring can be implemented so that if trends demonstrate that one treatment was significantly better or worse than the other with respect to any study endpoints (mortality, morbidity, side effects) it would be the responsibility of a special committee to determine whether the study should be terminated.

Long term follow-up is important in clinical trials since patients sometimes do not adhere to the originally assigned therapy.

3. Analysis and publication phase

Some issues of relevance to the analysis of data from randomized clinical trials include: baseline comparability of treatment groups, selection of prognostic factors, methods for evaluating treatment differences, non-adherence to assigned therapy, and post-stratification. Survival analysis is often the method of choice. A major consideration is how to analyze data when 1) persons are discovered, after randomization, who do not meet entry criteria, 2) withdrawals, 3) noncompliant subjects, 4) subjects who switch treatments. Exclusion of any groups will "undo" the pure randomization scheme and could result in a biased estimate of effect.

Advantages and disadvantages of RCTs:

Advantages	Disadvantages
1. Prospective	1. Contrived situation
2. Randomization	2. Human behavior may be difficult to control
3. Clear temporal sequence	3. Ethical constraints
4. Best evidence for causation	4. Exclusions may limit generalizability
	5. Expensive in time, personnel, facilities, and budget

Case-control studies

Of the three remaining classic epidemiologic study designs – cross-sectional, cohort or follow-up and case-control – the case-control is the least straightforward. We will therefore devote the following section to examining the "anatomy" and "physiology" of case-control studies.

Definition of a case-control study

A study that starts with the identification of persons with the disease (or other outcome variable) of interest, and a suitable control (comparison, reference) group of persons without the disease.

The relationship of an attribute to the disease is examined by comparing the diseased and nondiseased with regard to how frequently the attribute is present or, if quantitative, the levels of the attribute in each of the groups. — Last JM, A dictionary of epidemiology. 2nd edition, NY, Oxford, 1988

Synonyms: case comparison study, case compeer study, case history study, case referent study, retrospective study

Defining characteristic

Subjects are selected on the basis of the outcome variable.

Key advantages

- Statistically efficient for rare conditions
- Logistically efficient for prolonged induction or latency diseases
- Can examine many exposures in one study
- Ethical - cannot affect onset of disease

Basic procedure

1. Identify cases, determine their characteristics - cases estimate the prevalence of the exposure in people who get the disease.
2. Select controls (noncases), determine their characteristics - controls estimate the prevalence of the exposure in people who have not developed the disease.
3. Compare the characteristics of cases with characteristics of noncases.
4. Draw inferences about the underlying processes that led to differences in characteristics of cases and controls. Odds ratio ($OR = \text{odds of exposure in cases} / \text{odds of exposure in controls}$) estimates the incidence density ratio ($IDR = \text{rate of disease in exposed persons} / \text{rate of disease in unexposed persons}$). For rare disease, IDR closely approximates cumulative incidence ratio (CIR, RR) of the disease for that exposure.

$$\text{OR}_e \text{ (Exposure odds ratio)} = \frac{(a / m_1) / (b / m_1)}{(c / m_0) / (d / m_0)} = \frac{ad}{bc}$$

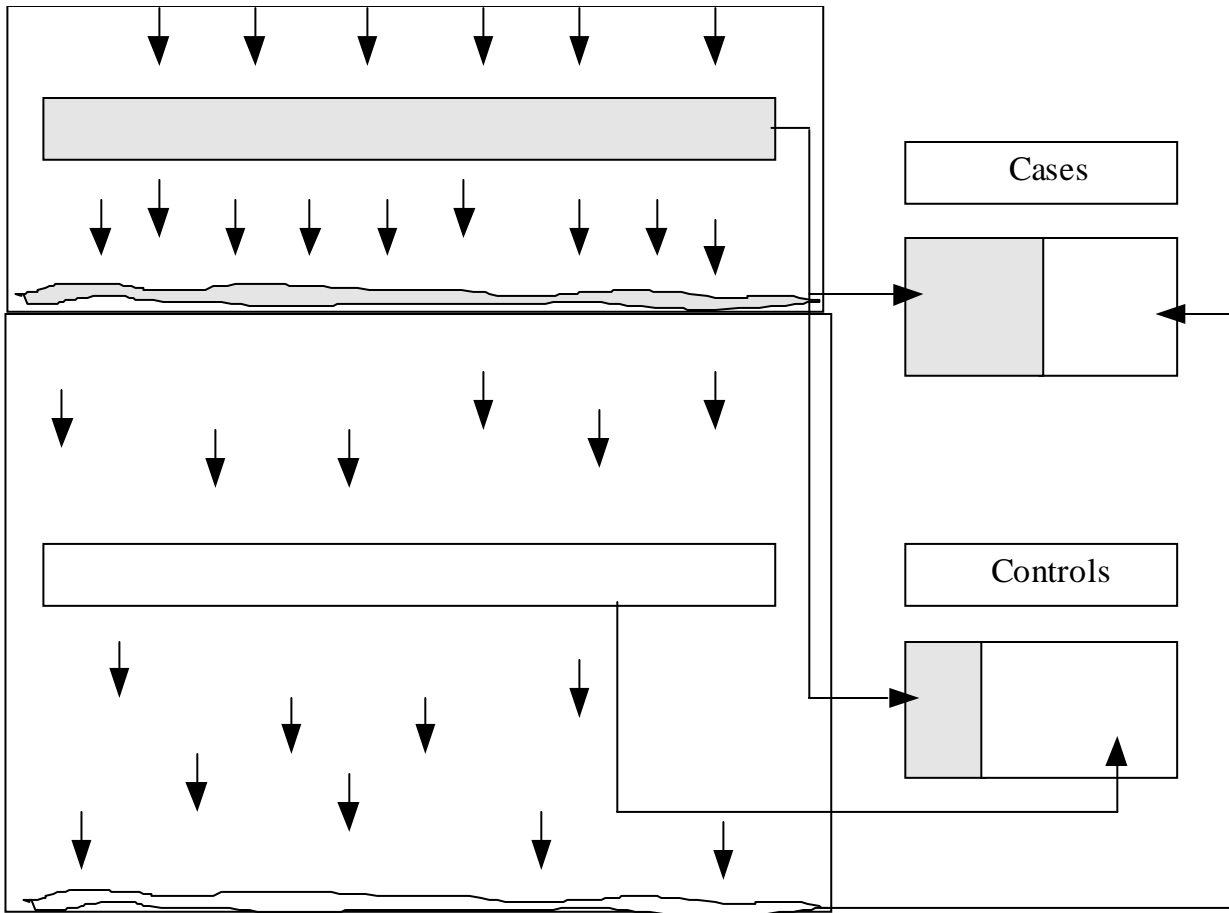
estimates IDR and CIR (rare disease)

If we wish to obtain an estimate of the incidence density ratio (or the relative risk) for endometrial cancer with respect to estrogen use, we can use the proportion or prevalence of estrogen use in the endometrial cancer cases to compute the odds of estrogen use among women who develop endometrial cancer [$p_{\text{estrogen}|\text{case}}/(1-p_{\text{estrogen}|\text{case}})$] and the proportion or prevalence of estrogen use in the controls to compute the odds of estrogen use in the population [$p_{\text{estrogen}|\text{noncase}}/(1-p_{\text{estrogen}|\text{noncase}})$]. The odds ratio for exposure is then the ratio of these two odds, and gives us the estimate of the relative risk (since endometrial cancer is a rare disease) and, if we have selected our cases and controls appropriately, of the incidence density ratio.

Rationale for the odds ratio

1. The cases provide an estimate of the prevalence of the exposure in people who get the disease.
2. The number of exposed cases (and therefore the proportion or prevalence of exposure among cases) reflects the rate of disease in exposed people in the population. The number of unexposed cases reflects the rate of disease in the unexposed population.
3. The odds of exposure in the cases (proportion exposed/proportion unexposed) therefore reflect the ratio of disease rates (or risks) in the population.
4. The controls provide an estimate of the prevalence of the exposure characteristic in the population from which the cases arose.
5. The odds of exposure in the controls (proportion exposed/proportion unexposed) reflect the odds of exposure in the population.
6. So the odds ratio (OR) [odds of exposure in cases/odds of exposure in controls] indicates the relative risk [incidence of disease in exposed persons/incidence of disease in unexposed persons].

The above rationale is presented to convey a "feel" for why the odds ratio from a case-control study conveys information about the strength of association between a disease and exposure.



Validity

The validity of a case-control study requires that:

- Cases in the study adequately represent the relevant cases (the population of cases about whom inferences are to be made) with respect to the variables of interest (notably, prevalence of exposure). This depends upon whether the cases available do in fact reflect the rates of disease in exposed and unexposed individuals undistorted by differential manifestation, detection, or short-term survival (e.g., selective survival, access to care, detection bias);
- Controls accurately reflect the exposure proportions in the study base (the source population for the cases). For example, hospitalized controls may overrepresent exposures associated with hospitalization for other conditions.

Both of these requirements, especially the latter, can be difficult to ensure. Therefore, case-control studies are regarded as highly susceptible to bias from problems with the:

Identification of cases

- Reliance on medical care system
- Often miss subclinical cases (detection bias?)
- Can miss rapidly fatal cases (selectively?)

Selection of controls

- Selection of controls can determine the study results
- Which controls are appropriate is often not obvious
- Trade-off between sampling and data collection
- Hospitalized controls, community controls, dead controls
- Controls may be reluctant to cooperate

Measurement of exposure for cases and controls

- Reliance on recall or records (differential?)
- Effect of disease on exposure assessment
- Effect of disease on exposure (confounding by indication)

There is also the thorny problem of establishing temporality, i.e., did the exposure precede the disease?

Interpretability of the odds ratio

Why does the OR from the cases and controls we have assembled estimate anything in the population? Consider what the cells in the table below represent. Assume that the cases were selected as newly occurring cases of endometrial cancer over a period of time in a defined

population and that the controls were selected at the same time as the cases from among women in that population (this is called "density sampling of controls").

Endometrial cancer	Estrogen		Total	
	Yes	No		
Case	a	b	m ₁	(a + b)
Control	c	d	m ₀	(c + d)
Total	n ₁	n ₀	n	
	(a + c)	(b + d)		

If this situation, the cases would be all (or some fraction f_1 of) cases of endometrial cancer in the population. If the incidence rate of endometrial cancer is ID and the amount of population-time is N women-years, then:

$$m_1 = (f_1)(ID)(N)$$

[f_1 is included only for purposes of generality — if all cases are included, then $f_1=1$ and can be ignored.]

Cases among women taking estrogen (cell "a") would be:

$$a = (f_1)(ID_1)(N_1)$$

where ID_1 and N_1 are the incidence rate and population-time, respectively, for women taking estrogen.

Similarly, cases among women not taking estrogen (cell "b") would be:

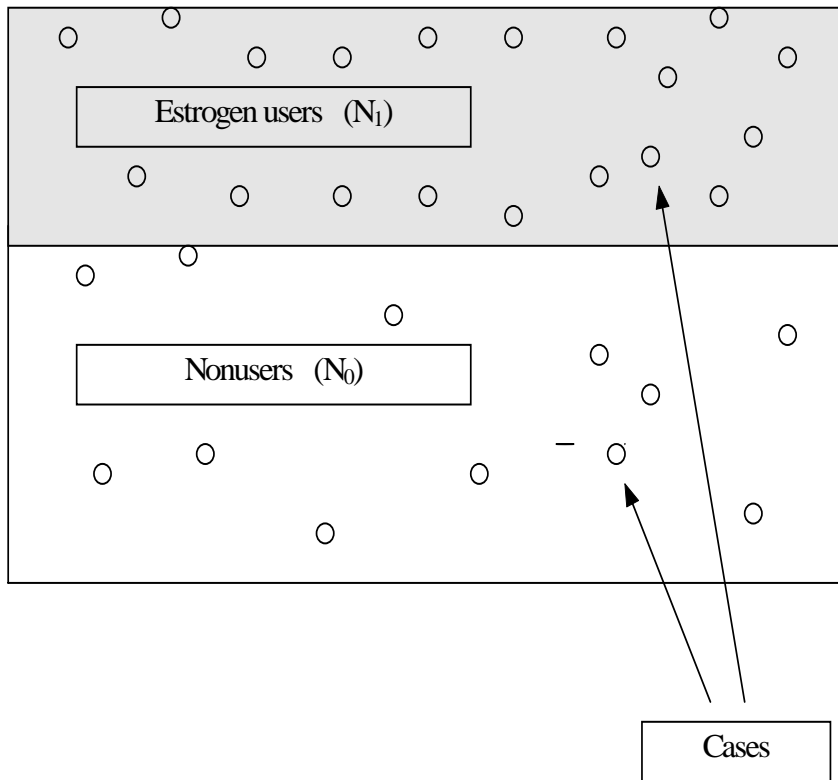
$$b = (f_1)(ID_0)(N_0)$$

with ID_0 and N_0 applying to women not taking estrogen.

Note: Whether N_1 and N_0 represent women-years of estrogen use or women-years in estrogen users (i.e., are person-years for a women after she stops taking estrogen counted as exposed or unexposed) would depend upon whether the estrogen effect endures after the drug is discontinued.

We now see where the cases have come from. What about the controls? The control group is typically, though not necessarily, chosen in some fixed ratio to the number of cases, such as two controls per case.

Since the strategic advantage of the case-control method is that we do not need to enroll the entire population from which the cases arise, the number m_0 of controls will be some small fraction f_0 of the noncases in the population. If we have 200 cases and have decided to select 400 controls, then f_0 would be 400 divided by the population size or the amount of population-time. f_0 is included to demonstrate the link between the case-control study and the population from which the cases arise, also referred to as the study base. In actual practice we establish the number of cases and controls required to meet certain sample size (statistical power) objectives; the sampling fraction f_0 is what results from the number of controls we seek. Since we do not – must not – choose our controls separately from each exposure group, the number of exposed (c) and unexposed (d) controls will be determined by the amount of population-time in each exposure category:



New cases among exposed	=	$ID_1 N_1$
New cases among unexposed	=	$ID_0 N_0$
Exposure odds in cases	=	
Exposure odds in noncases	\approx	N_1 / N_0

$c = f_0N_1 =$ number of exposed noncases

$d = f_0N_0 =$ number of unexposed noncases

If the N's represent population-time, rather than simply population size, f_0 reflects sampling over time as well as over people.

The key point about f_0 is that for the control group to provide a valid estimate of the relative sizes of exposed and unexposed population-time, f_0 must be the same for both exposed controls (c) and unexposed controls (d).

The above discussion can be summarized in a revised 2 x 2 table:

	Exposed	Unexposed
Cases	$f_1ID_1N_1$	$f_1ID_0N_0$
Controls	f_0N_1	f_0N_0

With this background, we are ready to see how the OR can estimate the IDR:

$$OR = \frac{ad}{bc} = \frac{(f_1ID_1N_1)(f_0N_0)}{(f_1ID_0N_0)(f_0N_1)} = \frac{ID_1}{ID_0} = IDR$$

Numerical example

Assume a stable, dynamic population of 4 million women age 40 years or older, in which 1,000 incident cases of endometrial cancer occur each year (i.e., 1,000 cases/4 million women-years).

Increase rate to 2,500 cases/ 100,000 wy?

Assume:

- $N_1 = 1$ million women-years (1,000,000 wy or 1×10^6 wy) of estrogen use
- $N_0 = 3$ million women-years (3×10^6 wy) of unexposed person-time
- ID_1 (incidence density in exposed) = 40×10^{-5} /year (40/100,000 wy)
- ID_0 (incidence density in unexposed population) = 20×10^{-5} /year, so that the IDR is 2.0

In the 1×10^6 exposed women-years, there would be 400 cases.

In the 3×10^6 unexposed women-years, there would be 600 cases.

Of the 1,000 cases, 400 are exposed and 600 are unexposed. The prevalence of exposure among cases is $400/(400+600) = 40\%$; the exposure odds in cases would be $.40/.60 = 0.67$.

The expected prevalence of exposure in an unbiased sample of noncases would be, since the disease is so rare, $N_1/(N_1+N_0) = (1 \times 10^6) / (1 \times 10^6 + 3 \times 10^6) = 0.25$; the exposure odds among noncases would be $0.25/0.75 = 0.33$.

The exposure odds ratio (OR) would therefore be:

$$OR = (.40/.60)/(.25/.75) = 0.67/.33 = 2.0$$

A case-control study that recruited (randomly) 200 cases and 400 controls ($f_1 = 200/1,000 = 0.2$; $f_0 = 400/4,000,000 = 1/10,000$ or 10^{-4}) would be expected to have the following results.

**Expected Results from Hypothetical, Population-based
Case-Control Study of Endometrial Cancer and Estrogen**

	Exposed	Unexposed	Total
Cases	80	120	200
Controls	100	300	400
Total	180	420	600

$$OR_e = \frac{80 \times 300}{120 \times 100} = 2.0$$

It is apparent that given the rarity of the disease it makes no practical difference here whether the prevalence of exposure in the source population from which the cases emanate (i.e., the study base) is estimated from the total population or only from those without the disease.

Identifying the study base

Disease and other types of events occur in populations. Case-control studies provide a window into the process of disease occurrence in a population, without the necessity of studying the entire population. Thus, case-control studies are best understood by considering what is happening in the population (the study base) and by analyzing the relationship between it and the case-control study.

But how do we identify the study base? The study base or source population consists of those people who would have been available to be counted as cases had they developed the disease or

experienced the event under study. Thus, the source population must be at risk for the disease and for being selected as cases if they were to develop it. Moreover, the relevant exposure measure for both cases and for the source population is the time during which the disease was initiated or promoted in the cases. Identifying the relevant period in time can be an issue for a disease with a lengthy induction and/or latent period, such as most cancers, if the disease is common and/or the population or its exposure distribution is undergoing substantial change.)

The first step in identifying the study base is generally based on geography or membership. Thus, for cancer cases from a state with a tumor registry, the study base is the state (or a portion of the state if only cases from a certain portion of the state are being studied). For cases detected in a managed health care organization, the study base is its membership. If identification of cases is being made through hospitals, then the study base is the population from whom people would go to that hospital if they developed the disease. This last situation can be complicated by factors such as the extent to which some people go to hospitals not covered in the study and whether the disease is one which does not always lead to hospitalization.

An important next step is to identify that subset of the population that is truly at risk for the disease (and its detection). For endometrial cancer, obviously the study base does not include men. Does the study base include hysterectomized women? Certainly not, since women without a uterus obviously cannot develop endometrial cancer – though if the hysterectomy was recent, a woman could be part of the study base for cases detected prior to that time. (Also, there may be the potential for selective depletion of endometrial cancer susceptibles, but we will not consider that possibility here.)

Selecting a control group representative of the study base

At least as problematic as identifying the study base is coming up with a way to obtain a control group that will faithfully represent it. One obvious choice, which is now much more common than in earlier decades, is to carry out a random sample survey of the study base as it exists at the time of the study.

This approach is most likely to be valid if:

- an accurate sampling frame exists or is constructed
- a representative sample is drawn and adequately executed
- response rates are high and data are of adequate quality (high rate of accuracy)

Controls recruited from hospitals and other noncommunity-wide sources are nevertheless of interest because the cost and logistical challenges are often not as great, greater cooperation may be obtained, and data quality may be better than that from the general population. However, when controls are obtained from sources other than a random sample survey, validity depends upon whether these controls have the same exposure distribution as the study base. For example, selecting controls from friends of the cases ("friend controls") can lead to bias because people tend to choose friends because of shared interests, perspectives, affiliations, and so on which are often associated with exposures. Thus, the proportion of many exposures in friend controls will be more

similar to that in the case group than in the study base as a whole. The use of friend controls is an example of "over-matching".

What about if some subsets of the study base are at much higher risk than others, due to, for example, genetic factors or simultaneous exposures? If the difference in risk is great, then both the case group and study base should be demarcated on that risk factor, and separate (stratified) analyses carried out.

Variants in the basic case-control design

There are several ways in which the case-control study design can be implemented.

- Incident versus prevalent cases: Case-control studies can use only new cases (*incident cases*) of the disease, thereby avoiding some of the sources of bias inherent in the use of *prevalent cases* (e.g., influence of survival/duration of the condition), or they can use prevalent cases.
- *Defined population* or nesting: Case-control studies can be carried out in a geographically-defined population, e.g., a state where a cancer register provides notification of all incident cases from a known denominator population, or in a cohort that has been followed (e.g., an occupational group). Having a defined population offers further advantages (such as availability of an identified universe for selection of controls, knowledge of the denominator from which migration has occurred, measurement of key variables prior to the disease). A case-control study within an identified cohort is sometimes termed a "nested case-control" study. (Rothman and Greenland regard nearly all case-control studies as nested in their source population.)
- Eligible controls: Although classically, the controls in a case-control study were noncases, in some designs people who later develop the disease can still serve as controls.

Types of control groups - case-control, case-cohort

Whether the OR in a case-control study estimates the IDR or the CIR depends upon the study design, particularly in relation to the sampling of controls (see Greenland 1987). Even if the disease is not rare, if it has an extended risk period (so that incidence density would generally be preferred) and the controls are obtained through "density sampling" (selected at the same time as matched cases), then the OR from the case-control study estimates the IDR (see below; ambitious students are referred to Sander Greenland and Duncan C. Thomas, "On the need for the rare disease assumption in case-control studies", *Am J Epidemiol* 1982; 116:547-53 and references therein).

The controls in a case-control study can be selected from among (a) persons who have not developed the disease by the end of the period of case ascertainment (prevalence controls), (b) persons who have not developed the disease at the time each case occurs - such controls are usually matched in time to the cases (density sampling), or (c) persons at risk to become a case at the outset of case ascertainment.

These controls may be selected before or after case ascertainment. Rodrigues and Kirkwood (1990) call the three types of controls, respectively, "exclusive", "concurrent", and "inclusive". The traditional approach is method (a), "exclusive" controls. With this method, only people who remain free of the disease to the end of case ascertainment are accepted as controls. The odds ratio in this situation estimates the incidence (i.e., risk) odds ratio in the cohort from which the cases arose. For a rare disease, this incidence odds ratio estimates the CIR.

In the second sampling scheme (density or concurrent sampling [method (b)], a participant can be selected as a control at a given point even if that participant later develops the disease. With this approach, the odds ratio computation estimates the relative rate (IDR) on the assumption that the IDR does not change during the follow-up period (assuming matching of controls to cases by time) (see Greenland and Thomas, 1982 and Rodrigues and Kirkwood, 1990). This study design has been referred to as a "density case-control study" (Hogue et al., 1983 referred to this design as a "case-exposure study"; however, Rodrigues and Kirkwood (1990) use that term for the third design [method (c)]). If a participant selected as a control later develops the disease, then that participant is also counted as a case; his/her data are used both as a case and as a control (his/her data appear in both categories).

The third design [method (c)] has been called "case-base" and "case-cohort" (also "case-exposure" – see Rodrigues and Kirkwood for citations). When such a case-control study is carried out within a fixed cohort, the odds ratio estimates the risk ratio with no rare disease assumption.

Rodrigues and Kirkwood show that the three ratio measures of association – CIR, IDR, and OR – can each be expressed so that its numerator is the odds of exposure in cases. Thus, all that differs are the denominators, and the three different approaches to sampling controls provide estimates for the respective denominators.

$$a. \quad OR_e = \text{Exposure odds ratio} = \frac{\text{Odds of exposure given disease}}{\text{Odds of exposure given non-diseased}} = \frac{a / b}{c / d}$$

(c / d is the odds of exposure in non-cases [never-cases at end of ascertainment period])

$$b. \quad IDR = \frac{ID_1}{ID_0} = \frac{a / py_1}{b / py_0} = \frac{a / b}{py_1 / py_0}$$

(py₁ / py₀ is ratio of exposed to unexposed person-years, from density sampling)

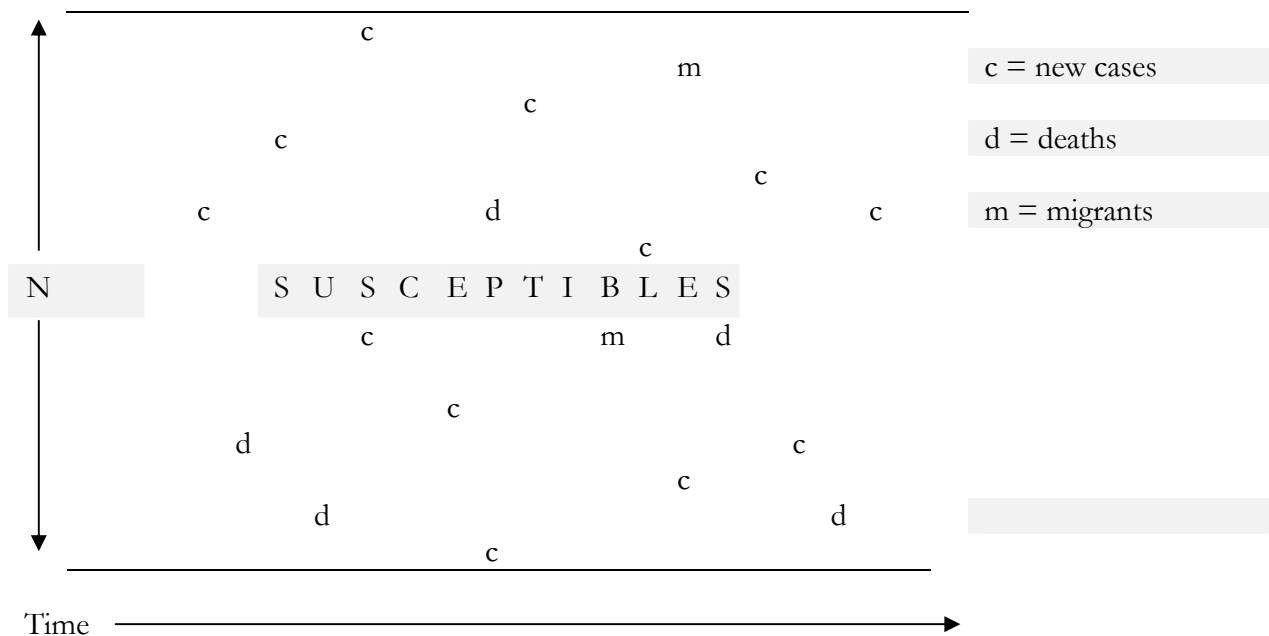
$$c. \quad CIR = \frac{CI_1}{CI_0} = \frac{a / n_1}{b / n_0} = \frac{a / b}{n_1 / n_0}$$

(n_1 / n_0 is the odds of exposure in the source population for the cases at the start of the follow-up)

where "a" = exposed cases, "b" = unexposed cases, and "n" and "py" represent persons and person-years for exposed (subscript 1) and unexposed (subscript 0).

A modern perspective

In general, design issues in a case-control study are best understood by considering how the issues would be dealt with in a randomized clinical trial (Feinstein, 1985) or a cohort study (Rothman and Greenland). In fact, students of epidemiology (including those of us on the other side of the podium) might have an easier time if the terms cohort study and case-control study had never been introduced, but rather the various approaches of studying disease presence and occurrence in a population classified in regard to the "windows" they provide into the development of the disease in the population.



The above diagram depicts a population of size N followed over time interval t . Suppose N_0 are susceptible (to a specific outcome) and that a surveillance system exists to detect cases (c's) of various diseases or events. For the moment, let us focus on a particular disease, and assume that M cases develop during the follow-up period shown. We will also focus on a particular exposure, to

which N_1 of the population are exposed, leaving N_0 unexposed. We will designate the total population-time in the exposed group as N_{1t} and that in the unexposed group N_{0t} . The population distribution of disease and exposure are summarized in the following table.

	Exposure		
	Yes	No	Total
Cases	A	B	M_1
People	N_1	N_0	N
Incidence proportion	A/N_1	B/N_0	M_1/N
Incidence proportion difference		$(A/N_1) - (B/N_0)$	
Incidence proportion ratio		$(A/N_1) / (B/N_0)$	
Person-time	N_{1t}	N_{0t}	N_t
Incidence rate	$A/(N_{1t})$	$B/(N_{0t})$	$M_1/(N_t)$
Incidence rate difference		$(A/N_{1t}) - (B/N_{0t})$	
Incidence rate ratio		$(A/N_{1t}) / (B/N_{0t})$	

We can estimate any of the measures in this table with data from appropriately selected random samples and good historical information. For example, if we choose a random sample (n) from the original susceptible population (N), the ratio of exposed persons in the sample (n_1) to unexposed persons in the sample (n_0) estimates (N_1/N_0) the odds of exposure in the original population. If we then choose a random sample (of size m_1) of the M_1 cases (or obtain data from all M_1 cases), then ratio of cases in the sample who were exposed at the beginning of the period (a) to unexposed cases in the sample (b) estimates the odds of exposure in cases. By including only cases who were present in the population at the start of the period, we can then estimate the incidence proportion ratio $[(A/N_1)/(B/N_0)]$ as the ratio of the estimated odds of exposure in cases (a/b) divided by the estimated odds of exposure in the susceptible population at the start of the follow-up period (n_1/n_0). This estimate will be accurate if we have representative samples, accurate assessment of baseline exposure, and no loss to follow-up from outmigration or deaths. If in addition we know N, the size of the original susceptible population, then we can also estimate N_1 and N_0 as, respectively, $(n/N)n_1$ and $(n/N)n_0$, thereby allowing us to estimate incidence proportions and the incidence proportion difference. With this design we can estimate incidence density proportion ratios for any diseases for which a surveillance system (possibly our own) is available and any exposures for which we can obtain baseline data. Note that no rare disease assumption is involved in the above estimates.

If duration of follow-up time is important, we need to estimate the ratio of exposed and unexposed susceptible follow-up time. We can do this by sampling the susceptible population over time,

instead of at baseline, in such a way that the probability of selecting a person is proportional to the amount of time he/she is susceptible ("density sampling"). One method for doing this is "risk-set" sampling, in which a susceptible person is sampled at the same date that each case occurs. The ratio of exposed to unexposed persons sampled in this way estimates N_{1t}/N_{0t} , which we can use to estimate the incidence rate ratio.

Finally, if we choose to sample susceptibles at the end of the follow-up period (Rothman and Greenland call this the "cumulative" design), then we can estimate the incidence odds ratio, which if the disease is rare will approximate the incidence rate ratio and the incidence proportion ratio. See Rothman and Greenland, chapter 7.

Bibliography

Hennekens and Buring. Rothman - *Modern Epidemiology*, Chapters 3-4, 6. Rothman and Greenland, Chapters 5-7. Lilienfeld and Lilienfeld - *Foundations of epidemiology*, Chapters 8-10; Mausner & Kramer - *Epidemiology: an introductory text*, Chapters 3 and 7-8. MacMahon & Pugh - *Epidemiology principles and methods*, Chapters 5 and 11-12. Fletcher, Fletcher, and Wagner - *Clinical epidemiology*, Chapters 5-7 (through page 118). Schlesselman. Case-control studies. Chapter 1. Kleinbaum, Kupper, Morgenstern - *Epidemiologic research*. Chapters 4-5.

Gray-Donald, Katherine; Michael S. Kramer. Causality inference in observational vs. experimental studies. *Am J Epidemiol* 1988; 127:885-892. See also correspondence from *Am J Epidemiol* 1989; 130:206-209.

"Multilogue" in the *Journal of Chronic Diseases / Journal of Clinical Epidemiology*:

Kramer, Michael S.; Jean-Francois Boivin. Toward an "unconfounded" classification of epidemiologic research design. *J Chron Dis* 1987; 40(7): 683-688.

J Clin Epidemiol 1988; 41(8):

Miettinen, Olli S. Striving to deconfound the fundamentals of epidemiologic study design, 709-713.

Greenland, Sander, Hal Morgenstern. Classification schemes for epidemiologic research designs, 715-716.

Kramer, Michael S.; Jean-Francois Boivin. The importance of directionality in epidemiologic research design, 717-718.

J Clin Epidemiol 1989; 42(6):

Feinstein, Alvan R. Epidemiologic analyses of causation: the unlearned scientific lessons of randomized trials. 481-490.

Miettinen, Olli S. The clinical trial as paradigm for epidemiologic research, 491-496.

Feinstein, Alvan R. Unlearned lessons from clinical trials: a duality of outlooks, 497-498.

Miettinen, Olli S. Unlearned lessons from clinical trials: a duality of outlooks, 499-502.

Szklo, Moyses. Design and conduct of epidemiologic studies. *Preventive Medicine* 1987; 16:142-149 (from Workshop on Guidelines to the Epidemiology of Weak Associations)

Ecologic studies

Cohen, Bernard L. Invited commentary: in defense of ecologic studies for testing a linear-no threshold theory. *Am J Epidemiol* 1994;139:765-768.

Greenland, Sander; James Robins. Invited commentary: ecologic studies – biases, misconceptions, and counterexamples. *Am J Epidemiol* 1994;139:747-60.

Greenland, Sander; James Robins. Accepting the limits of ecologic studies. *Am J Epidemiol* 1994;139:769-771.

Greenland, Sander. Interpretation and choice of effect measures in epidemiologic analyses. *Am J Epidemiol* 1987;125:761-768.

Koopman, James S.; Longini, Ira M. The ecological effects of individual exposures and nonlinear disease dynamics in populations. *AJPH* 1994;84:836-842.

Piantadosi, Steven. Invited commentary: ecologic biases. *Am J Epidemiol* 1994;139:761-764.

Poole, Charles. Ecologic analysis as outlook and method. Editorial. *AJPH* 1994;84(5):715-716)

Schwartz S. The fallacy of the ecological fallacy: the potential misuse of a concept and the consequences. *AJPH* May 1994;84:819-824.

Susser, Mervyn. The logic in ecological: I. The logic of analysis. *AJPH* 1994;84:825-829. II. The logic of design. *AJPH* May 1994;84:830-835.

Cohort studies:

Liddell FDK. The development of cohort studies in epidemiology: a review. *J Clin Epidemiol* 1988; 41(12):1217-1237.

Samet, Jonathan M., Alvaro Muñoz. Cohort studies. *Epidemiologic Reviews* 1998; 20(1).

Case-control studies:

Armenian, Hartoune K. Applications of the case-control method. *Epidemiologic Reviews* 1994;16(1):1-164.

Cornfield, Jerome. A method of estimating comparative rates from clinical data. *JNCI* 1951;11:1269-1275. (A classic)

Feinstein, Alvan R. Experimental requirements and scientific principles in case-control studies. Editorial. *J Chron Dis* 1985; 38:127-134.

Gordis, Leon. Should dead cases be matched to dead controls? *Am J Epidemiol* 1982; 115:1-5.

Greenberg RS, Ibrahim MA. The case-control study. In: *Textbook of Public Health*. Holland W, Detels R, and Knox EG (ed). NY, Oxford, 1983.

Greenland, Sander. Control-initiated case-control studies. *Int J Epidemiol* 1985; 14:130-134.

Greenland, Sander; Duncan C. Thomas. On the need for the rare disease assumption in case-control studies. *Am J Epidemiol* 1982; 116:547-53.

Greenland, Sander; Duncan C. Thomas, Hal Morgenstern. The rare-disease assumption revisited: a critique of "Estimators of relative risk for case-control studies". *Am J Epidemiol* 1986; 124:869-876.

Hogue, Carol J. R., David W. Gaylor, Kenneth F. Shulz. Estimators of relative risk for case-control studies. *Am J Epidemiol* 1983; 118:396-408.

Hogue, Carol J. R.; David W. Gaylor, Kenneth F. Schultz. The case-exposure study: a further explication and response to a critique. *Am J Epidemiol* 1986; 124:877-883.

Horwitz, RI, Feinstein AR. Methodologic standards and contradictory results in case-control research. *Am J Med* 66:556-564, 1979.

Horwitz, Ralph I. The experimental paradigm and observational studies of cause-effect relationships in clinical medicine. *J Chron Dis* 1987; 40:91-99.

Horwitz, Ralph I., and Alvan R. Feinstein. The application of therapeutic-trial principles to improve the design of epidemiologic research. *J Chron Dis* 1981; 34:575-583.

Kramer, Michael S.; Jean-Francois Boivin. Toward an "unconfounded" classification of epidemiologic research design. *J Chron Dis* 1987; 40:683-688.

Mantel, N. and Haenszel, W.: Statistical aspects of the analysis of data from retrospective studies of disease. *J National Cancer Institute* 22:719-748, 1959. Read 719-732 only [Read past page 732 at your own risk]. (Another classic)

McLaughlin, Joseph K.; William J. Blot, Eric S. Mehl, Jack S. Mandel. Problems in the use of dead controls in case-control studies. I. General results. *Am J Epidemiol* 1985; 121:131-139. II. Effect of excluding certain causes of death. *Am J Epidemiol* 1985; 122:485-94.

Miettinen, Olli. Estimability and estimation in case-referent studies. *Am J Epidemiol* 1976; 103: 226-235.

Newman, Stephen C. Odds ratio estimation in a steady-state population. *J Clin Epidemiol* 1988; 41:59-65 (advanced).

Pearce, Neil; Harvey Checkoway. Case-control studies using other diseases as controls: problems of excluding exposure-related diseases. *Am J Epidemiol* 1988; 127:851-6 (advanced).

Poole, Charles. Exposure opportunity in case-control studies. *Am J Epidemiol* 1986; 123:352-358.

Poole, Charles. Critical appraisal of the exposure-potential restriction rule. *Am J Epidemiol* 1987; 125:179-183.

Rodrigues L, Kirkwood BR. Case-control designs in the study of common diseases: updates on the demise of the rare disease assumption and the choice of sampling scheme for controls. *Intl J Epidemiol* 1990;19:206-213.

Schlesselman, James J.; Bruce V. Stadel. Exposure opportunity in epidemiologic studies. *Am J Epidemiol* 1987; 125:174-178.

Stavraky, Kathleen M.; Aileen Clarke. Hospital or population controls? An unanswered question. *J Chron Dis* 1983; 36:301-308, and responses on pages 309-316.

Wacholder, Sholom; Joseph K. McLaughlin, Debra T. Silverman, Jack S. Mandel. Selection of controls in case-control studies. I. Principles. *Am J Epidemiol* 1992;135(9):1019-1028

Wacholder, Sholom; Debra T. Silverman, Joseph K. McLaughlin, Jack S. Mandel. Selection of controls in case-control studies. II. Types of controls. *Am J Epidemiol* 1992;135(9):1029-1041

Wacholder, Sholom; Debra T. Silverman, Joseph K. McLaughlin, Jack S. Mandel. Selection of controls in case-control studies. III. Design options. *Am J Epidemiol* 1992;135(9):1042-1050

Wingo, Phyllis A.; Howard W. Ory, Peter M. Layde, Nancy C. Lee, and the Cancer and Steroid Hormone Study Group. The evaluation of the data collection process for a multicenter, population-based, case-control design. *Am J Epidemiol* 1988; 128:206-217.

Plus several articles and comments in *J Chron Dis* 1985; 38(7) and responses in 1986; 39(7):565-571.

Intervention trials:

Chalmers, Thomas C.; Paul Celano, Henry S. Sacks, Harry Smith, Jr. Bias in treatment assignment in controlled clinical trials. *N Engl J Med* 1983; 309:1358-61.

Farquhar, John W.; Stephen P. Fortmann, June A. Flora et al. Effects of communitywide education on cardiovascular disease risk factors: The Stanford Five-City Project. *JAMA* 1990;264:359-365.

Feinstein, A.R. *Clinical Epidemiology: The Architecture of Clinical Research*. Chapter 29, pp. 683-718.

Feinstein, Alvan R. Current problems and future challenges in randomized clinical trials. *Circulation* 1984; 70:767-774.

Friedman L.M., Furberg, C.D., DeMets, D.L. *Fundamentals of clinical trials*. Boston: John Wright PSG Inc., 1982.

Koepsell TD et al. Symposium on community trials: Invited commentary, 594-599 plus articles by Allan Donner, Stephen Fortmann, Sylvan Green. *Am J Epidemiol* 1995;142(6)

Lavori, Philip W.; Thomas A. Louis, John C. Bailar III, Marcia Polansky. Designs for experiments—parallel comparisons of treatments. *N Engl J Med* 1983; 309:1291-8.

Lilienfeld, Abraham M. *Ceteris Paribus: The evolution of the clinical trial*. *Bulletin of the History of Medicine* 56:1-18, 1982.

Luepker R, et al. Community education for cardiovascular disease prevention: Risk factor change in the Minnesota Heart Health Program. *Am J Public Health* 1994;84:1383-1393.

MacMahon and Pugh - *Epidemiology: principles and methods*, Chapter 13; Mausner and Kramer - *Epidemiology: an introductory text*, Chapter 8; Lilienfeld and Lilienfeld - *Fundamentals of epidemiology*, Chapters 10-11.

Moses, Lincoln E. The series of consecutive cases as a device for assessing outcomes of intervention. *N Engl J Med* 1984; 311:705-710

Peto, R., Pike, M.C., Armitage, P. et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *Br J Cancer* 34:585-612, 1976.

Peto, R., Pike, M.C., Armitage, P., et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. Analysis and examples. *Br J Cancer* 35:1-39, 1977.

Sacks, H., Chalmers, T.C., Smith, H. Randomized versus historical controls for clinical trials. *Amer J Med* 72:233-239, 1982.

Schwartz, D., Lellouch, J. Explanatory and pragmatic attitudes in therapeutic trials. *J Chron Dis* 20:637-648, 1967.

Simon, Richard and Robert E. Wittes. Methodologic guidelines for reports of clinical trials. Editorial. *Cancer Treatment Reports* 1985; 69:1-3.

Weinstein, Milton C. Allocation of subjects in medical experiments. *N Engl J Med* 291:1278-1285, 1974.