AACC | Better health through laboratory medicine.

PEARLS OF LABORATORY MEDICINE

Principles of Study Design and Analysis: Case-Control Studies

Julie E. Buring, ScD

CCTC Clinical Chemistry Trainee Council

Totality of Evidence

- Any research question may be addressed using a number of different research strategies.
- Choice based on scientific question, resources; but each has its strengths and limitations that must be clearly understood.
- No one epidemiologic study can answer a research question definitively.
- To assess our current status of knowledge, need to consider the totality of evidence.

Totality of Evidence

- Basic Research – how? possible mechanism; laboratory or animal evidence
- Epidemiologic Studies – in humans
  - Descriptive – who, what, when, where?
    - correlational or ecologic studies
    - case reports and case series
    - cross-sectional studies or surveys
  - Analytic – why?
    - observational
      - case-control
      - cohort
    - intervention studies
      - randomized clinical trials

Epidemiologic Design Strategies

DESCRIPTIVE STUDIES

- Who? What? Where? When?
- Correlational or ecologic study
- Case reports/series
- Cross-sectional study

ANALYTIC STUDIES

- Why? Search for factors associated with or predictive of outcome
- Observational study
    - case-control
    - cohort
- Intervention study
  e.g., randomized clinical trial

Analytic Studies

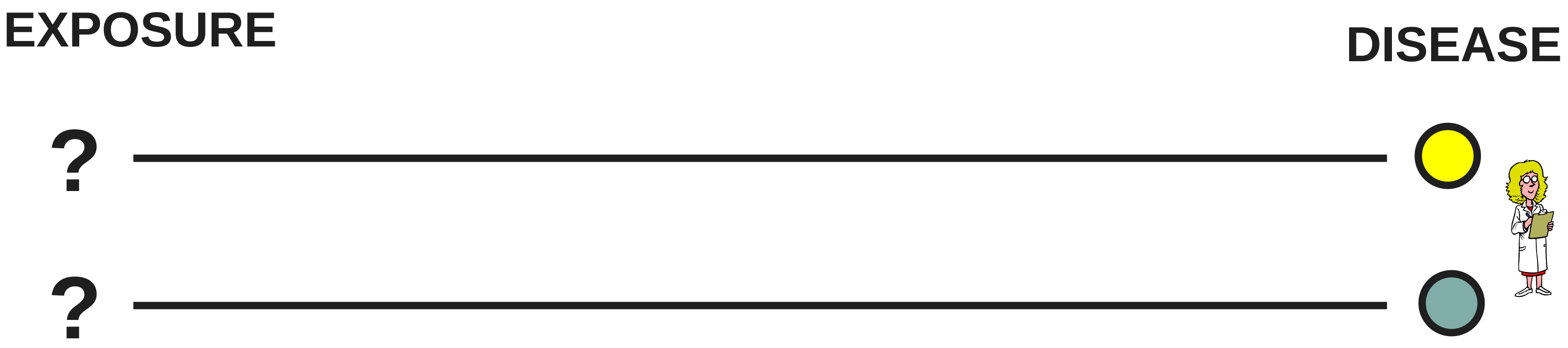
- Observational Studies (exposures are self-selected)
  - Case-control
    (initial selection on basis of disease status)
  - Cohort
    (initial selection on basis of exposure status)
- Intervention Studies (exposures are allocated by investigators)
  e.g., randomized clinical trial

Case-Control Study

A case-control study is a type of observational analytic epidemiologic study in which subjects are selected on the basis of whether they do (cases) or do not (controls) have a particular disease/outcome under study.

The groups are then compared with respect to the proportion having a history of an exposure or characteristic of interest.

Case-Control Study: Observational study, with selection into study on basis of disease status



○ PRESENT

● ABSENT

Basis on which groups are selected at beginning of study.

INVESTIGATOR

Case-Control Study: Example

QUESTION: **Is there an association between a baby's sleeping position (prone vs. back) and risk of Sudden Infant Death Syndrome (SIDS)?**

CASES: **All deaths from SIDS in infants aged 7 to 364 days in two regions of the UK from February 1993 through January 1994, reported through a communication network of professional organizations who report all sudden unexpected deaths.**

CONTROLS: **The next two younger and the next two older babies born in the same hospital within two weeks of the age of the index baby, who did not die of SIDS.**

Case-Control Study: Example

QUESTION: Do young women (less than 40) who regularly participate in physical exercise activities during their reproductive years have a reduced risk of breast cancer?

CASES: 545 women (aged 40 or younger at diagnosis) newly diagnosed with in-situ or invasive breast cancer between 7/1/83 and 1/1/89 identified by the population-based cancer registry for Los Angeles County.

CONTROLS: One neighborhood control individually matched to each case by date of birth (within 36 months), race, and parity (nulliparous vs parous), based on a predefined walk pattern for the neighborhood where the case lived at time of diagnosis.

Strengths of a Case-Control Study

- Approach began relatively recently, as diseases of interest shifted from acute to chronic public health problems.
- Solution to the logistic difficulties of studying diseases of long latent periods (long periods between adequate exposure and development of outcome).
- Efficient design with respect to time and money, since outcome has already occurred.

Strengths of a Case-Control Study

- Since select on basis of diseased status, can identify adequate numbers of diseased and nondiseased people.
- Thus, ideal design when outcome is rare; don’t need to follow large numbers of people in order to get sufficient numbers who subsequently develop particular outcome.

Strengths of a Case-Control Study

- Allows ideal for evaluation of multiple exposures or risk factors for a single disease outcome, as well as interrelationships among these factors.
- Can be used to test hypotheses, or in the absence of *a priori* hypotheses, explore a wide range of exposures ("fishing expedition") to test in subsequent studies. Particularly useful in early stages of knowledge about a disease or outcome.

Limitations of a Case-Control Study

- Major problem is susceptibility for bias, since both exposure and disease have already occurred when participants enter the study. When assessing exposure status, disease status already known.
- Potential for selection bias (differential selection of either cases or controls into study on basis of exposure status) as well as observation bias (differential reporting/recording of exposure between study groups based on disease status).

Limitations of a Case-Control Study

- **Temporal sequence of exposure and outcome may be difficult to establish. Logic can be considered going "backward" – going temporally from effect (disease) to cause (antecedent exposure). To reflect this, used to say "retrospective study" as synonym for case-control. But not unusual logic in everyday life.**
- **Do have to worry about ability to both get accurate past exposure information and for right time period.**
- **Cannot calculate incidence disease rates or relative or attributable rates directly, but can estimate using odds ratio.**

Bottom Line

- These are not reasons not to do case-control studies – they are just reasons to design carefully.
- Bottom line – case-control studies work! Most common analytic epidemiologic study design in medical literature. Often first study design used in analytic epidemiology – optimal approach when new condition, or when need to conserve money or time.

Sources of Cases

- Selection of cases not usually the difficult part.
- Need to define a disease or outcome of interest. Want as homogenous a disease entity as possible, since similar manifestations of disease have different etiologies (uterine/cervix; congenital malformations; issue of sample size).
- Strict diagnostic criteria for the disease, which are reproducible (WHO definition for myocardial infarction; stratify as definite, probably, possible).

Sources of Cases

- Hospital-based case-control study.
- Treated at medical facilities (hospitals, HMO's, private practices) during specified period of time.
- Relatively easy and inexpensive to conduct.

Sources of Cases

- Population-based cases. Selecting all or random sample in a defined general population.
- Involves locating and obtaining data from all affected individuals or random sample from a defined population.
- Avoids bias from selective factors that leads to use of particular hospital.
- If complete ascertainment, allows description of entire picture in community; allows direct computation of rates.
- But big logistic and cost considerations, so not frequently done.

Sources of Cases

- Other sources: disease registries; special surveys.
- Incident or prevalent cases?: Advantage of prevalent cases increases sample size. But price paid in terms of interpretation of whether risk factors for development or survival. Use incident cases if possible, to separate out effect of duration.
- "Representative cases"? – Random sample of all cases in population, rather than 5 biggest hospitals? Will increase generalizability – but likely increase logistic difficulties. Validity first, generalizability second.

Sources of Controls

- **Most difficult and critical issue for the design and validity of case-control study.**
- **Necessary to allow evaluation of whether what is observed in cases is different than what would be expected based on comparable people without the disease.**
- **No control group optimal for all situations.**
- **Depending on source from which cases where chosen, controls must be selected to represent not the entire nondiseased population but the population of individuals who would have been identified and included as cases had they also developed the disease.**
- **Like the cases, may not be representative of general population, but crucial requirement that they be comparable to the source population of the cases, and any exclusions or restrictions made in the identification of cases apply equally to the controls and vice versa.**

Sources of Controls

- **If hospitalized cases, consider hospitalized controls.**

Advantages:

- **convenient, easily identified, readily available in sufficient numbers, relatively inexpensive, minimal effort.**
- **cases and controls are likely to be similar in their accuracy of recall, because both "sick", so minimizes recall bias. Same selection factors of coming to a hospital.**
- **generally high level of cooperation of subjects, compared to healthy individuals, minimizing non-response bias.**

Sources of Controls

- **Using hospitalized controls**

Disadvantages:

  - **Controls are ill. Differ from healthy individuals in ways associated with illness or hospitalization in general. May not represent the exposure distribution in the population from which cases derived.**
  - **Disease for which controls are hospitalized may be associated with risk factors under study (smoking and lung cancer – wouldn't use bronchitis, COPD; alcohol and CHD, wouldn't use car or sports accidents). How can be confident diagnostic groups chosen truly unrelated to factors under study.**
  - **Selection factors leading to hospitalization in a particular hospital for a particular disease may differ between cases and controls (referral patterns, primary vs. tertiary hospitals).**

Sources of Controls

- **When cases from hospital but hospital controls not scientifically adequate, or when cases come from general population, take controls from general population.**

Advantages:

- **generally ensures comparability – came from same source population.**

Disadvantages:

- **often difficult to enumerate all members of population as basis for selecting individuals (town lists in MA).**
- **difficult to gain cooperation for participation – time, motivation. Non-response ALWAYS greater than for hospitalized cases – major threat to validity.**
- **expensive and time consuming.**
- **quality of information – may not recall exposures with same degree of accuracy as cases.**

Sources of Controls: Special Methods

- Random digit dialing
  - May match on area code and dialing prefix
  - Problems
    - Households without phone (SES)
    - Households without landlines (young)
    - Probability of being home may be related to some exposures (income, SES, exercise patterns)
    - Answering machines – screening calls
    - Increasing cell phone use – no lists of numbers

Sources of Controls: Special Methods

- Friends, relatives, spouses, sibs, neighbor controls
  - More likely to be cooperative than general population.
  - Degree of control of important confounding factors related to ethnic background, SES, current or early environment.

Sources of Controls: Special Methods

- Friends, relatives, spouses, neighbor controls
  - But if the study factor itself is one for which family members and friends are likely to be similar to the cases, will make cases and controls artificially alike with respect to exposure, and will underestimate the true effect of the exposure (diet, smoking, exercise, pet ownership, social interactions).
    - The case identifies the control:
    - Because they are at low risk …
    - Because they are at high risk …
    - May elect to chose control based on exposure habits
- May lead to overmatching; requires matched analysis to take this into account.

Selection of Cases and Controls

Cases

- **Cases must be selected independently of exposure.**

Controls

- **Ideally, the controls are a direct random sample of the reference population from which the cases originated.**
- **Controls must be sampled independently of exposure.**

Issues in Selection of Controls

- How many control groups? One, two, more?
- Ideally, want single control group most comparable to cases.
- Multiple control groups indicated when concern that one selected group has a specific deficiency that could be overcome by the inclusion of another control group (alcohol, diet, coffee and CHD: who in hospital not related? Use hospital plus population control groups).

Issues in Selection of Controls

- How many controls per case (control-to-case ratio)? 1:1? 2:1? 100:1?
- When the number of available cases and controls is large and the cost of obtaining information from both groups is comparable, the optimal control-to-case ratio is 1:1.
- When the sample size of cases is limited, with only a small number being available for study, or when the cost of obtaining info is greater for cases or controls, the control-to-case ratio can be altered.
- As control-to-case ratio increases, power increases.
- But increased power levels off at 4:1, UNLESS data are available at very little extra cost or “free” (sometimes see 10:1). Power increases little after 4:1.

Ascertainment of Disease/Exposure Status

- Any potential source of information must be assessed in terms of ability to provide both accurate and comparable information for all study groups.
- Disease sources: death records, case registries (SEER for cancer), office records, hospital admission or discharges, pathology logs.
- Exposure sources: study subjects themselves (interview or mail q'aire; from a surrogate/proxy (mother for child, spouse for dead patient), from information recorded in medical records.

Ascertainment of Disease/Exposure Status

- Procedures used to obtain info must be as similar as possible for cases and controls: place and circumstances of interviews; blinding of abstractors; interviewers and patients unaware of specific hypotheses; trying to minimize observation bias for probing questions by interviewers.
- Obtaining records completed before occurrence of outcomes is especially valuable, since unlikely that accuracy or completeness of data at baseline will be dependent on whether subject later developed the disease (birth certificates for birth weight and gestational age for childhood cancer, prenatal x-ray from OB records for congenital malformation).

Ascertainment of Disease/Exposure Status

- Need to decide the basis on which a given individual should and will be considered "exposed".
- What part of person's exposure history is relevant to the etiology of disease – requires some understanding of the mechanisms of the disease process as well as likely latent period.
- Smoking/lung cancer – not amount currently smoked, but total duration of smoking. Smoking and MI – current is most relevant.
- If period too wide – "ever used" when "current" is what is relevant– then random misclassification, bias to the null.
- Can evaluate data from differing time windows of exposure, and gain info about period that appears most relevant.

Analysis of Case-Control Study

1. Set up data in 2x2 or rxc table.
2. Cannot directly calculate incidence measures of disease frequency (unless population-based case control study).
3. Estimate measures of association/difference

- Odds ratio = OR = $\frac{ad}{bc}$
- $AR_e\% = \frac{OR-1}{OR}$; PAR%

Role of Chance: Data-derived Hypotheses

- Most case-control studies test a small number of specific hypotheses.
- Most investigators also collect data on a multitude of potential risk factors, and conduct many comparisons.
- Important to distinguish between tests of hypotheses specified in advance (*a priori* hypotheses), and “fishing expeditions” in which associations emerge when data analyzed (data-derived hypotheses, *a posteriori* hypotheses).

Role of Chance: Data-derived Hypotheses

- **These must be interpreted with caution. Remember meaning of p-value: 1/20 comparisons will be statistically significant by chance alone if Ho true. Even if statistically significant, if in a hypothesis-formulating situation, still interpret as possibly due to chance.**
- **These formulated hypotheses can then be tested in studies specifically designed to do so.**

Special Type of Case-Control Study

- Nested case-control study within a cohort.
- Population is defined first (e.g., cohort study).
- Out of that population, cases develop over time and are identified; random sample of non-cases selected as control subjects.

Nested Case-Control Design

- Between 1987 and 1992, 10,786 women (ages 35–69 years) were recruited into a prospective study on breast cancer in Italy.
- At recruitment, urine was collected from all participants and stored at -80ºC. After an average of 5.5 years follow-up, 144 breast cancer cases were identified.
- Does estrogen metabolism affect breast cancer risk?
- Took 144 breast cancer cases; selected 576 controls from the cohort who had not developed breast cancer (matched on age, other variables). Conducted the assays on just these individuals.

Nested Case-Control Design

- Why did they do a nested case-control study?
- Efficiency (time and money).
- Assume the assay costs $25: entire cohort cost = $269,650. If did assay on all 144 cases and between 1–4 controls per case: $7,200–$18,000.
- Reduction of normal bias inherent in case-control study – information (q'aire plus blood) obtained prior to breast cancer diagnosis because the parent study is a cohort study.
- Very commonly done, especially in these tight financial times – ancillary case-control study nested into (grafted onto) existing parent cohort study (resource).

Summary: Case – Control Study

- Strengths
    - ideal for rare outcomes (inefficient for rare exposures – then turn to cohort study)
    - ideal if want to evaluate multiple exposures (risk factors) for a single outcome
    - efficient in terms of time and money
- Limitations
    - cannot calculate incidence rates (but can estimate relative measures)
    - potential for selection and observation bias (unless nested case-control)
    - difficulty in knowing appropriate time window for assessing exposure and getting accurate past exposure information

Thank you for participating in this *Clinical Chemistry* Trainee Council Webcast

Find our upcoming Webcasts and other Trainee Council information at www.traineecouncil.org

Follow us

f in