

University of Massachusetts Medical School  
**eScholarship@UMMS**

---

PEER Liberia Project

UMass Medical School Collaborations in Liberia

---


2018-11

## Foundations in Global Health Research Methodologies

Ann M. Moormann  
*University of Massachusetts Medical School*

Let us know how access to this document benefits you.

Follow this and additional works at: [https://escholarship.umassmed.edu/liberia\\_peer](https://escholarship.umassmed.edu/liberia_peer)

 Part of the [Family Medicine Commons](#), [Infectious Disease Commons](#), [International Public Health Commons](#), [Medical Education Commons](#), and the [Quantitative, Qualitative, Comparative, and Historical Methodologies Commons](#)

---

### Repository Citation

Moormann AM. (2018). Foundations in Global Health Research Methodologies. PEER Liberia Project. <https://doi.org/10.13028/yth4-hg49>. Retrieved from [https://escholarship.umassmed.edu/liberia\\_peer/3](https://escholarship.umassmed.edu/liberia_peer/3)

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in PEER Liberia Project by an authorized administrator of eScholarship@UMMS. For more information, please contact [Lisa.Palmer@umassmed.edu](mailto:Lisa.Palmer@umassmed.edu).

# Foundations in Global Health Research Methodologies

Feb 13, 2015

Ann Moormann, PhD, MPH

Associate Professor

Program in Molecular Medicine

And

Co-Director, Office of Global Health

# When I hear the word 'research'...

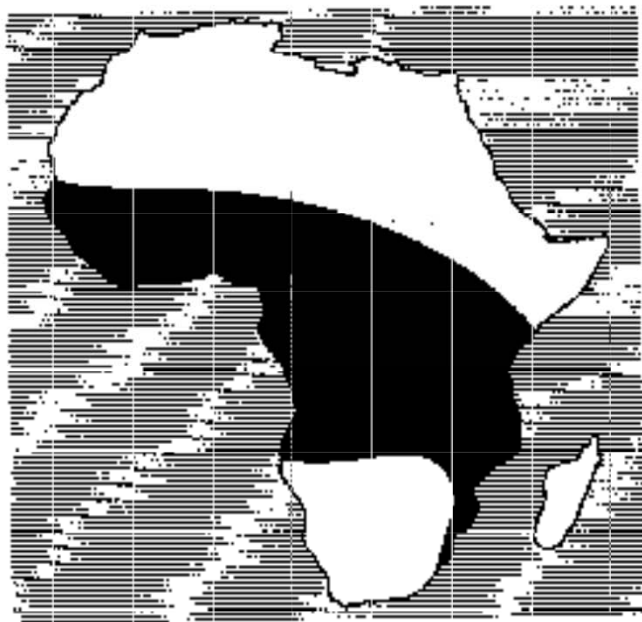
- \* I feel overwhelmed
  - \* I don't think it applies to me and what I want to do in my career – but it's fine if other people want to do it
- or**
- \* I get excited about learning something new
  - \* I want to get involved so I can figure out how to better care for my patients or prevent them from becoming ill

“How to did I get sucked into this research project??”



It starts with one patient....and then a few more...and collecting a little data...

Asking the question - Are other doctors seeing the same type of patients I am?



Lymphoma Belt:  
Geographic distribution  
mapped in 1962



Map of Rainfall:  
lined to high malaria transmission in 1967

# First described by Dr. Denis Burkitt in 1958 while working in Uganda

A COMBINED MEDICAL AND SURGICAL STAFF MEETING

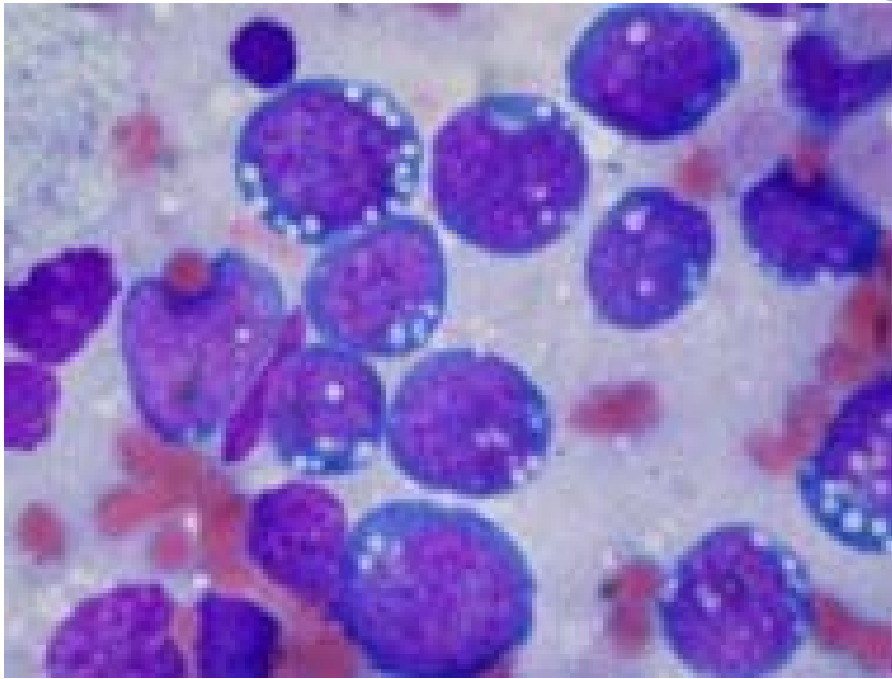
will be held

on Wednesday, 22nd March, 1961 at 5:15 p.m.

IN THE COURTAULD LECTURE THEATRE.

Mr. D.P. Burkitt from Makerere College,  
Uganda will talk on "The Commonest Children's  
Cancer in Tropical Africa. A Hitherto Unrecognised Syndrome".

Which led someone in the audience to ask  
another question...



Virus isolated from BL tumor  
in 1964  
by Tony Epstein and Yvonne  
Barr and Bert Achong

**Epstein-Barr Virus** was the  
first virus associated with a  
human cancer.

## 50 years later... we also know that



- \* Endemic Burkitt lymphoma is the most common pediatric cancer in equatorial Africa.
- \* Annual incidence 2-5 per 100,000 children.
- \* Peak incident age 5-9 years
- \* Sex ratio 1.5 males to females
- \* It's an extranodal monoclonal B cell tumor.
- \* High tumor proliferation index – 6 week from onset.
- \* Presentation: 33% Jaw, 48% abdomen, 11% jaw & abdomen, 11% eye, 3% CNS
- \* Survival rate ~70% using Chemotherapy

**There are over 10,000 publications on BL and over 30,000 publications on EBV since they were both first described by one physician 50 years ago.**



# Learning objectives

- \* What is epidemiology and translational research?
- \* What are standard study designs?
- \* How is data collected?
- \* How is data analyzed?
- \* When to seek IRB approval?
- \* How to find a research mentor?
- \* How is research funded?
- \* **Ultimate goals:** changes in health-care practices and public health policies.

# What is epidemiology?

- \* The study of the distribution and determinants of disease in human populations – and the application of this information to control or prevent diseases.
- \* The basic science of Public Health
- \* It's Quantitative – used to study the incidence and prevalence of disease within a population.
  - \* **Incidence** = rate at which people get a disease
  - \* **Prevalence** = number of people living with a disease
  - \* **Epidemic** = meaning 'upon or above' the 'people' – simply defined as when a disease occurrence is higher than expected based on recent experience.

# What does epidemiology do?

- \* Answers the following questions on a population-based level.
  - \* Who is getting this disease? (demographics)
  - \* Where do they live? (geographic)
  - \* When do they get this disease? (seasonal?)
  - \* How are they getting this disease? (risk factors and causes)
  - \* How do we stop this disease from affecting more people? (interventions)... translation of research findings to changes in clinical or public health practices

# Overview of Study Designs

## ***Descriptive Studies***

- \* Ecological
- \* Case report
- \* Case series
- \* Cross-sectional survey

## ***Analytic Studies***

- \* *Observational*
  - \* *Case-control*
  - \* *Cohort (retrospective and prospective)*
- \* *Experimental (Interventions)*
  - \* *Clinical Trials*

# Why Learn Study Designs?

Depending on the specific research question at hand, different study designs will help you to answer different questions:

- \* Randomized clinical trials (RCTs) for treatment response questions
- \* Cohort studies for prognosis questions
- \* Case-control and cohort studies for questions about etiology

# Why Learn Study Designs?

Immediately tells you several important things about the study (and its planning):

- \* How likely a study may be subject to bias compared to other types of studies (e.g., case-control studies are much more prone to bias than RCTs);
- \* What types of bias to look for in assessing the internal validity of the study (e.g., confounding bias is a problem in observational cohort studies but should not be as much a problem in an RCT); and
- \* Which measures of occurrence and measures of association are best used to report study results.

# Descriptive Studies

Descriptive studies reveal patterns of disease occurrence: person, place and time.

These studies provide general observations concerning the relationship of a disease to:

- \* basic demographic and social characteristics of the population under study (“person”);
- \* the geographic location of the study population (“place”); and
- \* the time of occurrence of the disease (“time”).

# Analytic Studies

- \* The goal of analytic studies are to determine whether certain characteristics of a population and their environment are associated with disease outcome or illness onset.
- \* These relationships may be ones of statistical association or not; they may be causal associations or not; and they may have indirect or direct causal associations.
- \* Investigators also often assess these relationships, whether statistically significant or not, as having clinical significance or not.



# Analytic Studies: Experimental vs Observational

- \* Analytic studies fall into 2 main categories: **observational** or **experimental**.
- \* **Experimental studies** (e.g., randomized controlled clinical trials) are designed such that investigators deliberately manipulate certain factors or exposures and, over time, observe the outcome. By controlling the experimental situation, investigators may conclude that the intervention or manipulation actually affected, or 'caused', a change in the outcome.
- \* Because of the difficulty of performing well-controlled experiments with human populations, where it's not always easy to change social or behavioral factors, and with the abundance of available observational data, observational studies are conducted much more frequently.

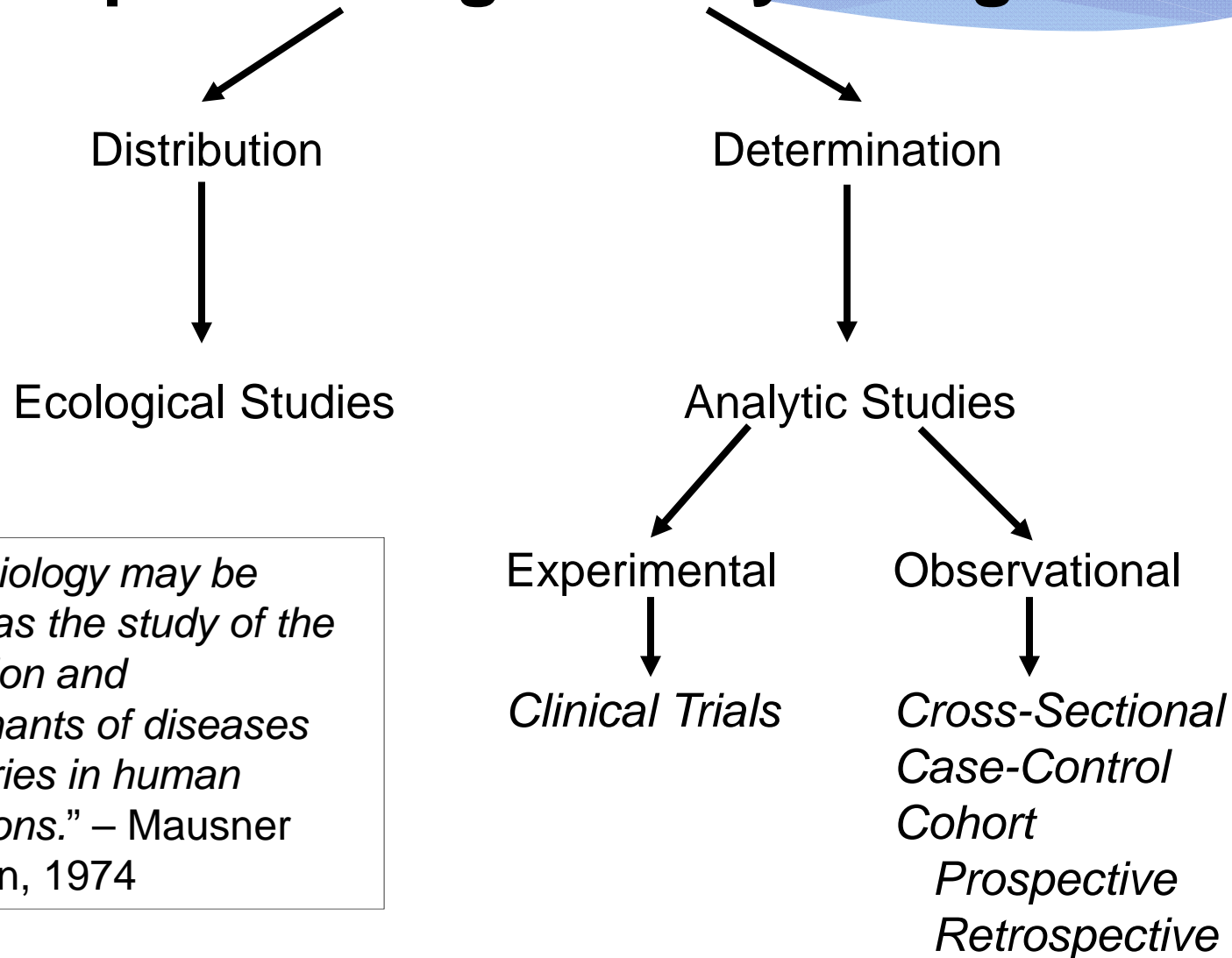
# Analytic Studies: Experimental vs Observational

- \* **Observational studies** (e.g., cross-sectional studies, case-control studies, and prospective / retrospective cohort studies) are designed whereby investigators don't manipulate the factors under study, but rather let nature take its course and observe the outcome.
- \* While observational studies are more representative and realistic than experimental studies, they are also more prone to bias.

# Analytic Studies: Experimental vs Observational

- \* Although incidence studies are usual preferable (RCTs and prospective cohorts), there is also an important role for prevalence studies (case-control and cross-sectional studies), both for practical reasons, and because such studies enable the assessment of the level of morbidity and the population “disease burden”.
- \* There are numerous advantages and disadvantages to the various study designs commonly used

# Epidemiologic Study Designs



*“Epidemiology may be defined as the study of the distribution and determinants of diseases and injuries in human populations.” – Mausner and Bahn, 1974*

# Ecological study

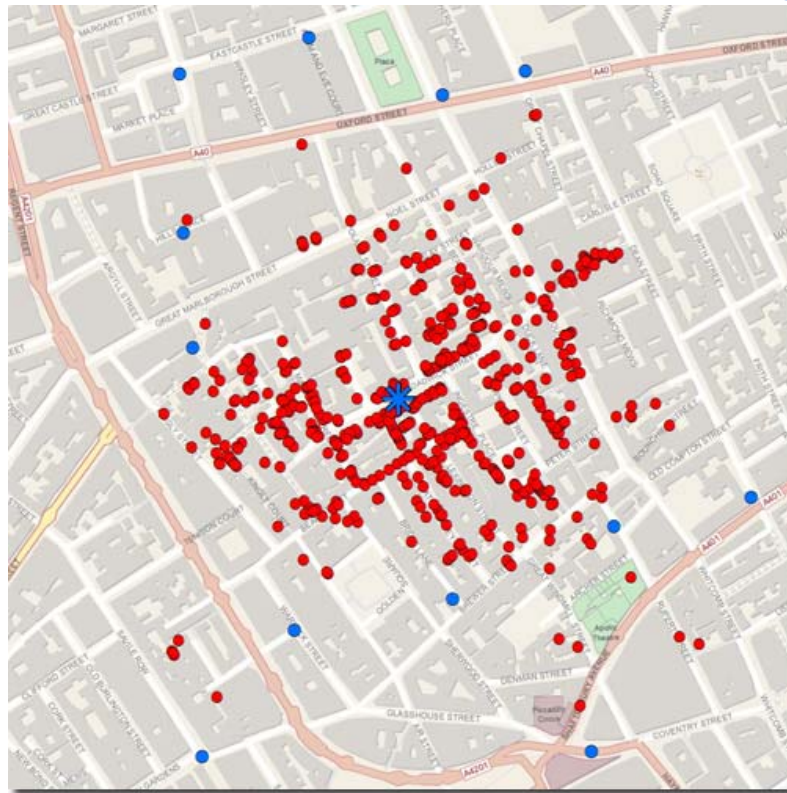
Studies that link risk-modifying factors for health based on populations defined either geographically or temporarily.

# The first ecological study

- \* The study by Dr. John Snow regarding a cholera outbreak in London is considered the first ecological study to solve a health issue.
- \* He used a map of deaths from cholera to determine that the source of the cholera was a pump on Broad Street. He had the pump handle removed in 1854 and people stopped dying there.
- \* It was only when Robert Koch discovered bacteria years later that the mechanism of cholera transmission was understood.



Location of water pumps and cholera deaths – led to removal of pump handle – and end to cholera outbreak in London



**CHOLERA**  
AND  
**WATER.**  
**BOARD OF WORKS**  
FOR THE LIMEHOUSE DISTRICT,  
Comprising Limehouse, Ratcliff, Shadwell,  
and Wapping.

The **INHABITANTS** of the District within  
which **CHOLERA IS PREVAILING**, are  
earnestly advised

**NOT TO DRINK ANY WATER**  
**WHICH HAS NOT**  
**PREVIOUSLY BEEN BOILED.**

Fresh Water ought to be Boiled every  
Morning for the day's use, and what  
remains of it ought to be thrown away  
at night. The Water ought not to stand  
where any kind of dirt can get into it,  
and great care ought to be given to see  
that Water Butts and Cisterns are free  
from dirt.

BY ORDER,  
**THOS. W. RATCLIFF,**  
CLERK OF THE BOARD.

Board Office, White Horse Street,  
7th August 1854.

# Ecological study

## Strengths

- \* Can be performed using data sets that are readily available to generate and test hypotheses.
- \* Studies can be performed relatively quickly and inexpensively.
- \* Can screen large numbers of people and examine many risk-modifying factors.
- \* Correlation coefficient provides some measure of the extent to which a risk factor is associated with a particular disease.

## Limitations

- \* Since measurements are for entire populations, and not specific individuals, it is impossible to establish that the presence of a risk factor is linked to health outcome of interest.
- \* Presence of a correlation does not imply a valid association.
- \* Subjected to ecological fallacy – findings for group may not apply to individuals within the group



# Bradford Hill Criteria for Causation

The list of the criteria is as follows:

1. **Strength:** A small association does not mean that there is not a causal effect, though the larger the association, the more likely that it is causal.
2. **Consistency:** Consistent findings observed by different persons in different places with different samples strengthens the likelihood of an effect.
3. **Specificity:** Causation is likely if a very specific population at a specific site and disease with no other likely explanation. The more specific an association between a factor and an effect is, the bigger the probability of a causal relationship.
4. **Temporality:** The effect has to occur after the cause (and if there is an expected delay between the cause and expected effect, then the effect must occur after that delay).
5. **Biological gradient:** Greater exposure should generally lead to greater incidence of the effect. However, in some cases, the mere presence of the factor can trigger the effect. In other cases, an inverse proportion is observed: greater exposure leads to lower incidence.
6. **Plausibility:** A plausible mechanism between cause and effect is helpful (but Hill noted that knowledge of the mechanism is limited by current knowledge).
7. **Coherence:** Coherence between epidemiological and laboratory findings increases the likelihood of an effect. However, Hill noted that "... lack of such [laboratory] evidence cannot nullify the epidemiological effect on associations".
8. **Experiment:** "Occasionally it is possible to appeal to experimental evidence".
9. **Analogy:** The effect of similar factors may be considered.

# Case Reports

Most case reports are on one of six topics:<sup>[1]</sup>

1. An unexpected association between diseases and symptoms.
2. An unexpected event in the course of observing or treating a patient.
3. Findings that shed new light on the possible pathogenesis of a disease or an adverse effect.
4. Unique or rare features of a disease.
5. Unique therapeutic approaches.
6. A positional or quantitative variation of the anatomical structures.

# Case Reports (publishable)

## \* **Strengths**

- \* Useful to call attention to a unique or mysterious case.
- \* Can report both adverse or beneficial effects
- \* High sensitivity for detecting novelty
- \* Can provide new ideas in medicine
- \* Rapid short communication (only one patient)

## \* **Limitations**

- \* Anecdotal evidence

# Case series or clinical series

- Tracks patients with known exposure given similar treatment or examines medical records to find link between exposure and outcome
- Can be prospective or retrospective
- Smaller sample size
- Can be consecutive or non-consecutive
- Can be confounded by selection bias and lack of an appropriate comparison group – which limits making causal link to disease.

## “Interesting” Case Reports

Between October 1980 and May 1981, 5 young men, all active homosexuals, were treated for biopsy-confirmed *Pneumocystis carinii* pneumonia at 3 different hospitals in Los Angeles. Two of the patients died. All 5 patients had laboratory-confirmed previous or current CMV infection and candidal mucosal infection.

Patient 1: A previously healthy 33-year old man developed *P. carinii* pneumonia and oral mucosal candidiasis in March 1981 after a 2 month history of fever associated with elevated liver enzymes, leukopenia, and CMV viremia. The serum complement-fixation CMV titer in October 1980 was 256; in May 1981 it was 32. The patient's condition deteriorated despite courses of treatment with pentamidine and acyclovir. He died May 3. Postmortem examination showed residual *P. carinii* and CMV pneumonia, but no evidence of neoplasia.

MMWR: June 5, 1981

# Cross-Sectional Survey

A cross-sectional survey is a study in which simultaneous assessments of outcomes, descriptive features, and potential predictors are made (“snap-shot” moment)

## Cross Sectional Surveys

### **Strengths**

Low cost method to monitor health status and health care needs of populations over time.

Determines annual **prevalence** of disease in a population.

Sometimes useful for suggesting possible associations between risk factors and disease.

Useful when disease is relatively common – can be used to monitor impact of public health intervention

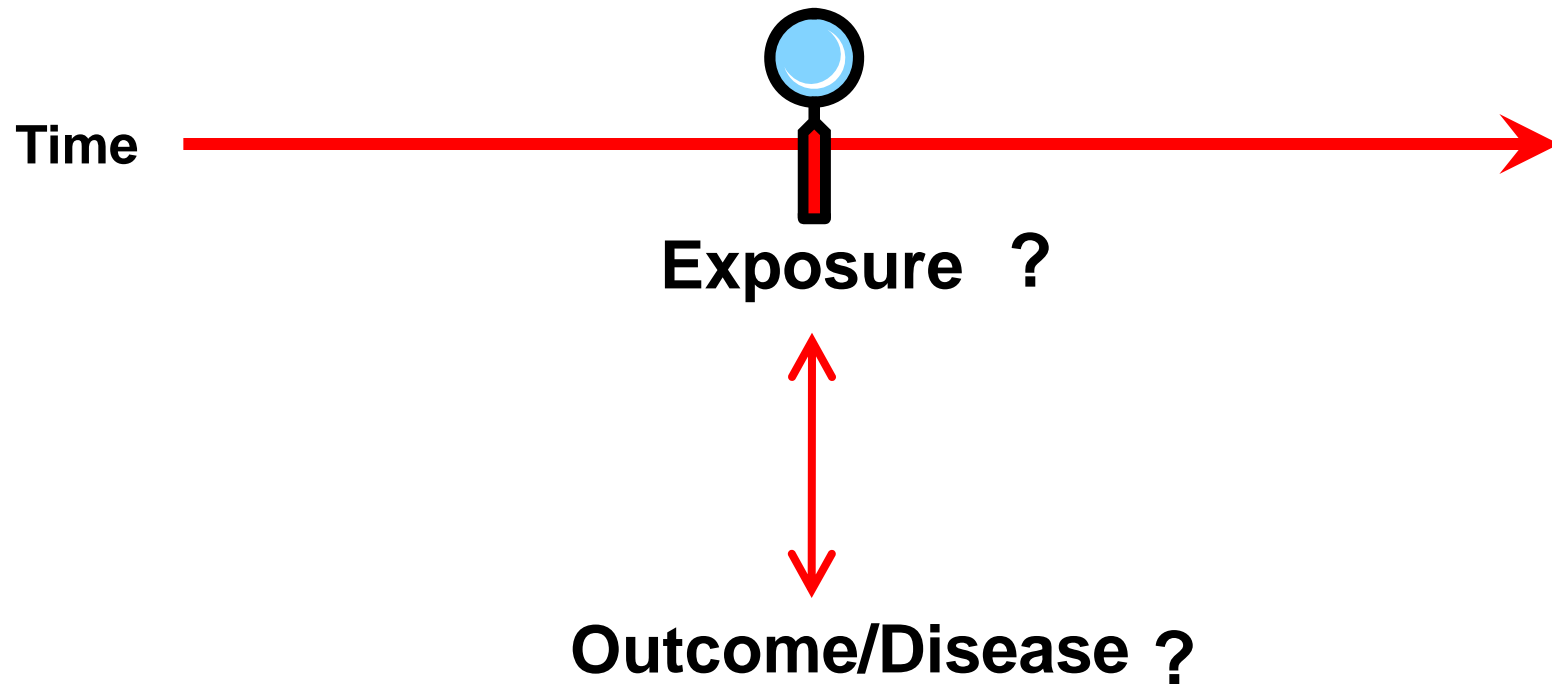
### **Limitations**

Since the presence of a risk factor and disease are assessed at the same point in time, the temporal relation between the risk factor and disease is unclear.

Subject to recall bias for past exposures

# Cross Sectional Surveys

## Non-Directionality



- Do people with the exposure all have the disease?
- Are people without the exposure all healthy?
- How many exposed people do not have the disease?
- How many people with the disease did not have the exposure?



# Outcomes Research

- \* Clearly define your outcome before you start your study.
- \* Is it rare or frequent in the target population?
- \* How will you measure it?
- \* How long should your study be to achieve the outcome?
- \* Are you sure you are not mis-classifying people?

# Inclusion versus exclusion criteria

- \* **Inclusion criteria** determine who will be recruited for enrollment in the study based on the study design, disease and exposures of interest. Clearly defined before you begin your study.
- \* **Exclusion criteria** determines who will be eliminated from analysis or factors that would disqualify them from the study. There should be a clear rationale that helps further refine the study population and avoid potential confounders.

# Cross-Sectional Survey examples

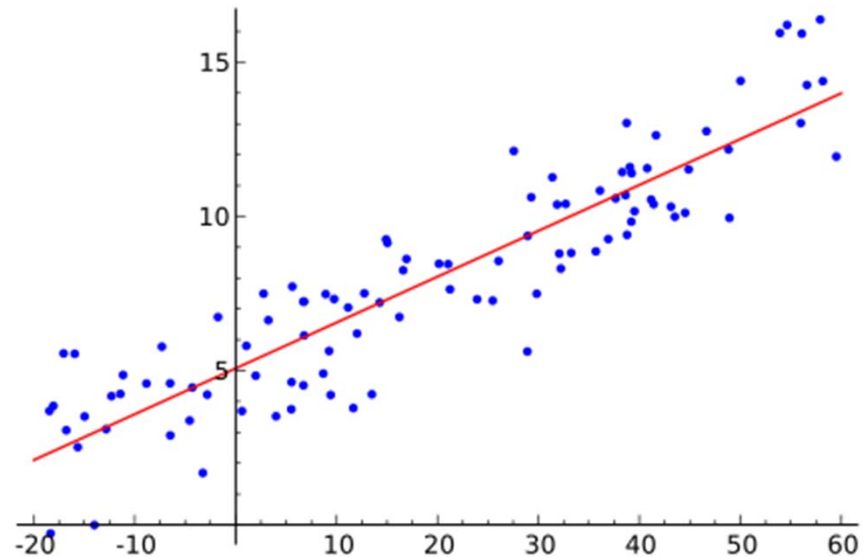
## Examples:

1. Is there an association between obesity and television watching?
2. Is there an association between being a construction worker and lung cancer?

# Data analysis for cross-sectional studies

To determine positive or inverse associations between exposure and outcomes

- Correlations
- Linear regression methods



# Case-control study

The comparison between two groups to make a link between an exposure and a disease:

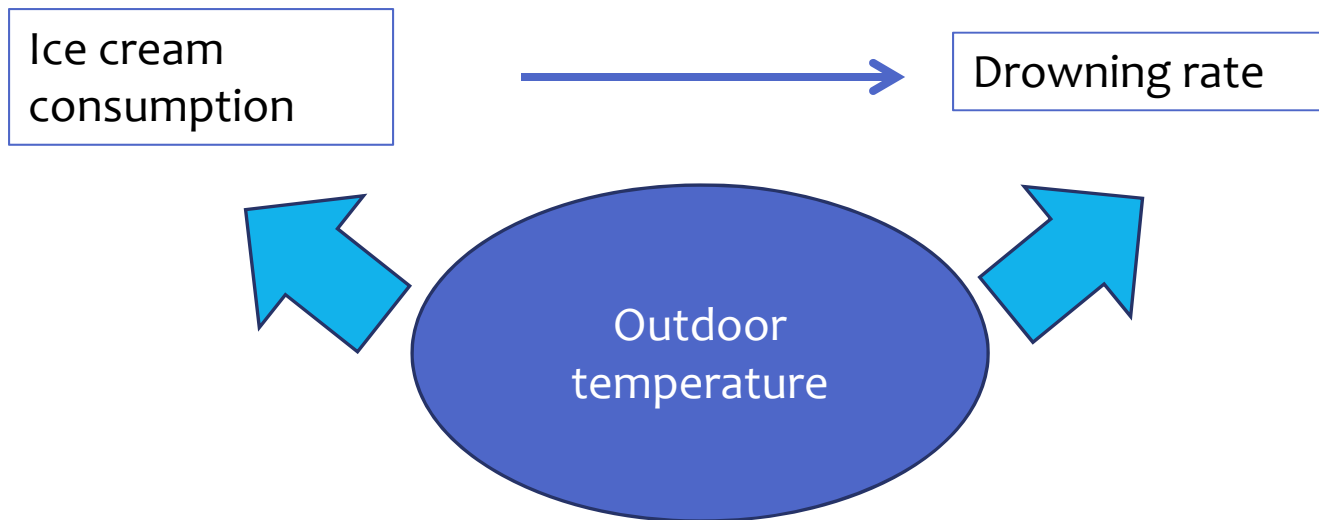
- **cases** have the disease outcome of interest
- **controls** do not have the disease

They are both asked about the same exposures and associations are measured using Odds Ratios

# Effect modifier versus confounder

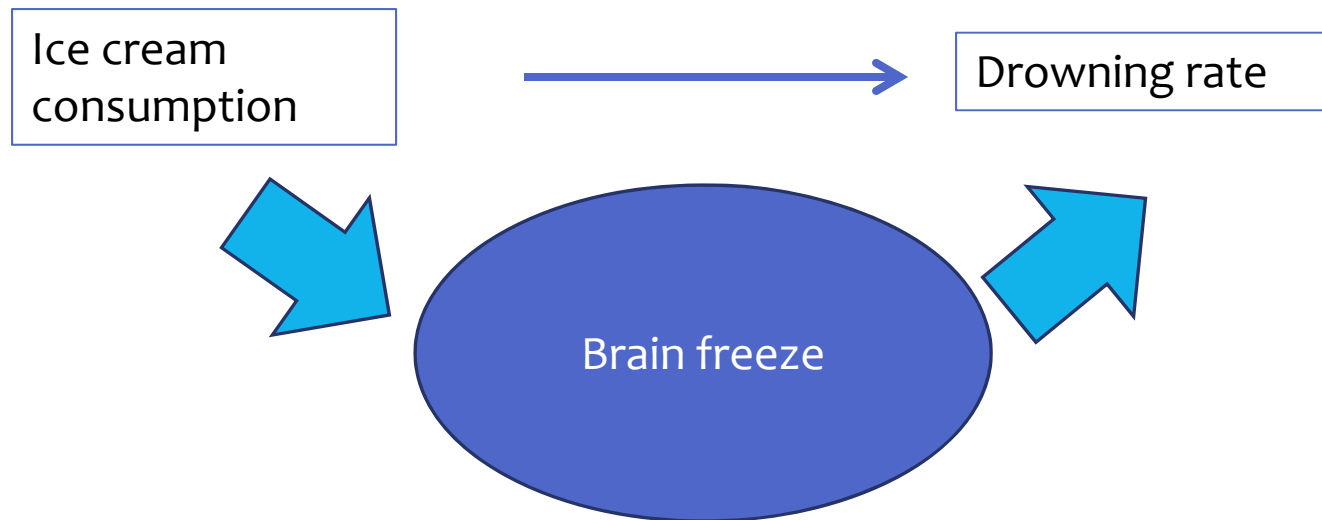
- ❖ **Confounding** means to ‘confuse’ – defined as a characteristic that is associated with both the risk factor and the outcome– but it is not in the causal pathway. In this instance you are comparing otherwise dissimilar groups and the outcome is distorted by the presence of another (unmeasured, lurking) variable.
- ❖ **Effect modifier (interaction)** means to influence (modify) the relationship between the risk factor and the outcome. The effect is real but the magnitude of the effect is different for different groups.

# Example of confounding



The outdoor temperature is associated with ice cream consumption and drowning rate but it is not in the causal pathway.

# Example of effect modification



Cramping is associated with ice cream consumption and thereby influences the drowning rate. The effect modification is only related to the outcome, not the exposure.



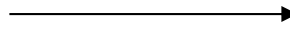
# The presence of confounding or effect modification can lead to inaccurate results

- \* Control for confounding by ‘adjusting’ for that variable. Failure to control for confounding can lead to over- or under-estimation of the true effect. Control for confounding by stratification of data or logistic regression analysis.
- \* Control for effect modification by comparing effects across groups. Stratify by group and see if the same cause-effect relationship exists – is there an interaction? The RR or OR differs according to the different levels.
- \* Minimizing either starts with a good study design and anticipating potential confounders or effect modifiers.

# Observational studies

## ***Case-control study***

Patients with disease, or with pre-specified outcome, and comparison group without disease

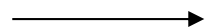


Look back to determine exposure to possible risk factors or causes

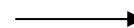
## ***Cohort study***

- To determine causes of disease

Population free from disease



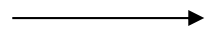
Patients followed



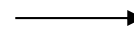
Incidence of new cases

- To determine natural history of disease

Population of patients with known disease



Patients followed



Incidence of pre-specified outcomes (e.g., death, decline in functional status)

# Case-control studies

- \* Strengths

- \* Useful when disease is rare (like cancer)
- \* Relatively inexpensive and quick

- \* Limitations

- \* Cannot calculate disease prevalence
- \* Not helpful for rare exposures
- \* Can only study one outcome (predetermined)
- \* Subject to selection bias and recall bias

# Case-control study: Classic example

You decide to host Thanksgiving dinner and invited 40 of your fellows and residents (and a few medical student). You ended up working late and didn't have time to properly thaw the turkey so you cooked it at a higher temperature and quickly made all the trimmings. Everything tasted great and was ready by 2 pm but some people came later because they just got off service. Then because they were watching the football game no one thought to put the food away – but that was ok because a few people ate again later on that evening.

# Case-control study: Classic example

Then a week later.. Much to your horror, you find out that some of the people who attended your dinner party came down with gastroenteritis....so to clear your name (and your cooking) you decide to conduct an epidemiology outbreak investigation...

# What evidence would you need to collect?

- \* Number of sick (number well).
- \* When did they get sick? (days after suspect meal)
- \* What did each person eat?
- \* What time did they eat?

# Results

- \* Of the 40 people who attended the dinner, 25 got ill (62.5% attack rate)
- \* The dinner was on Sat and most people got ill on Wed (average incubation period 3 days).

<b>Food</b>	<b>Number sick</b>	<b>Number well</b>	<b>Total people</b>	<b>Attack rate</b>
Turkey	24	16	40	60%
mashed potoates & gravy	26	14	40	65%
Green beans	20	20	40	50%
pumpkin pie	12	15	27	44%

# Interpretation

- \* Which pathogen do you suspect?
- \* Which food was contaminated?
- \* Some people ate the suspect food but did not get sick – what are some possible explanations?
  - \* Recall bias – didn't remember exactly what they ate
  - \* Dose-response – maybe smaller, less contaminated portion
  - \* **But** your study was limited to questions about your dinner party.. they later all confessed that they ate at the hospital cafeteria before coming to your house because they heard you were not a very good cook.



# Incident Versus Prevalent Cases

**Incidence** is the number of new cases within a time period (those newly diagnosed)

**Prevalence** is the number of existing cases  
may be low if high mortality  
may be high if low mortality

# Selection of Controls in a Case-Control Study

- Multiple hospital controls
- Community or sibling controls
- Matching by selected characteristic
  - Individual
  - Group

# Hospital Controls

Controls are selected from hospital patients with illnesses other than the disease of interest.

## ***Strengths***

- Easily accessible and tend to be more cooperative than population-based controls.
- Hospital-based studies are much less expensive and time-consuming than population-based studies.

## ***Limitations***

- Not likely to be representative of source population that produced the cases.
- Hospital-based controls are ill and exposure of interest may be a determinant of the control illness as well as the disease of interest. A real association of exposure with disease of interest would likely be missed.

# Community (population-based) Controls

The ideal control group should be representative of population from which the cases are derived (source population).

In population-based studies, controls are selected from the community. Methods used to select controls include random telephone dialing, friends or neighborhood, and DMV listings.

- \* **Strengths:** cases and controls come from same source population, so they are similar in many, unmeasured ways.
- \* **Limitations:** difficult to obtain population lists and to identify and enroll healthy study participants. Recall bias greater in controls than cases.

# Calculating the Odds Ratio for case control studies

*Initially Select*

*Then measure past exposure to suspect risk factor(s)*

Risk Factor	Cases (disease (+))	Controls (disease (-))
Exposed (+)	a	b
Exposed (-)	c	d

## *Measures of Association*

- Differences in % exposed:  $\frac{a}{a + c}$  versus  $\frac{b}{b + d}$
- Odds Ratio:  $ad/bc$

## Case-Control Study Considerations

- How dependable is information obtained from the past and have data been collected in a reliable manner?
- Is recall bias operating? What attempts have authors made to assess effect of this potential bias?
- Are other biases evident? Are there inequalities in information gathering, sampling, or observation between comparison groups?

# Temporal Direction of Study Designs

**Cohort study**

**Exposure**



**Outcome**

**Case-control study**

**Exposure**



**Outcome**

**Cross-sectional study**

**Exposure**



**Outcome**



**Time**

## Cohort Studies

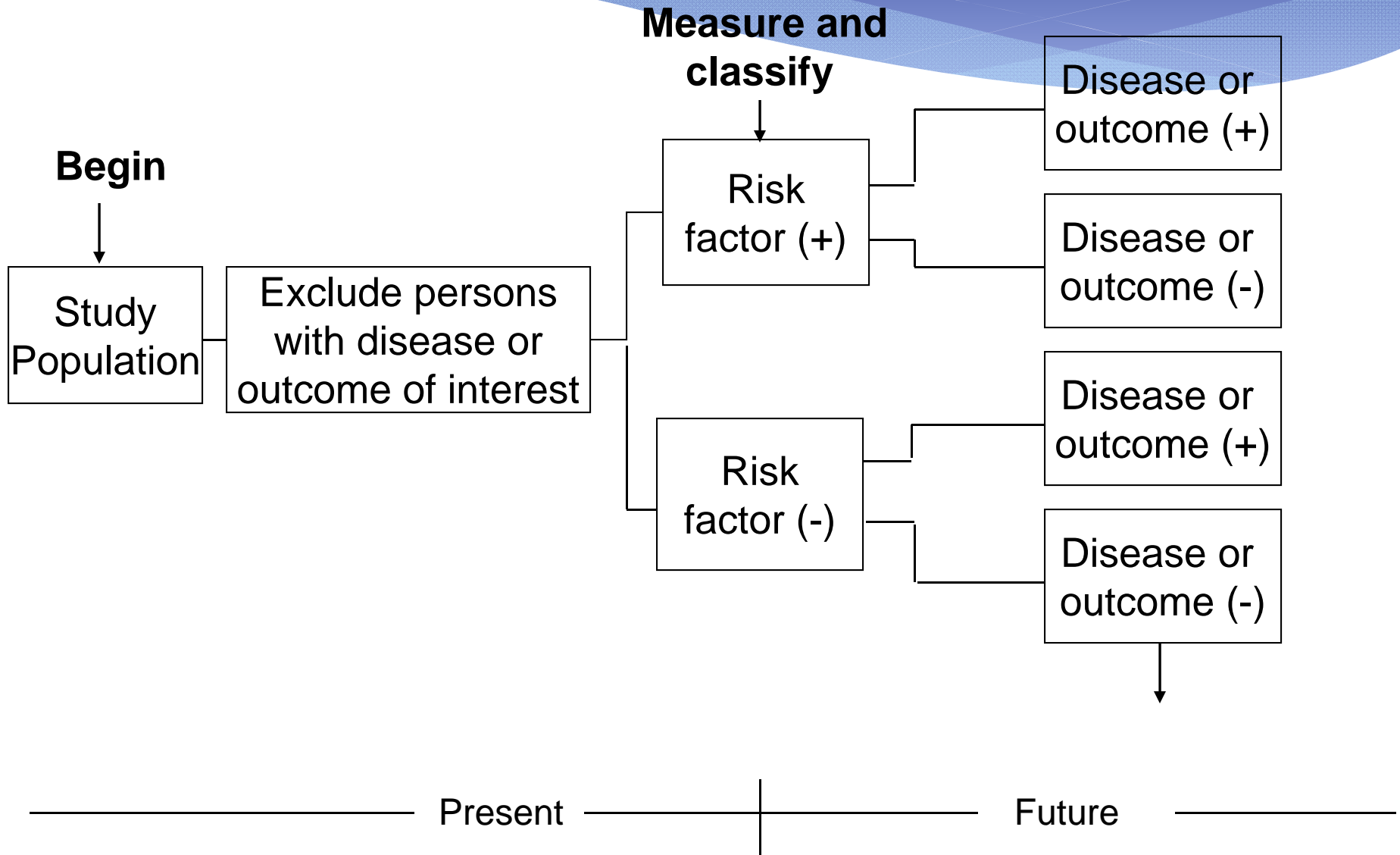
“Cohort” has its origin in the Latin *cohors*.

This phrase refers to warriors and the notion of a group of persons proceeding together in time.

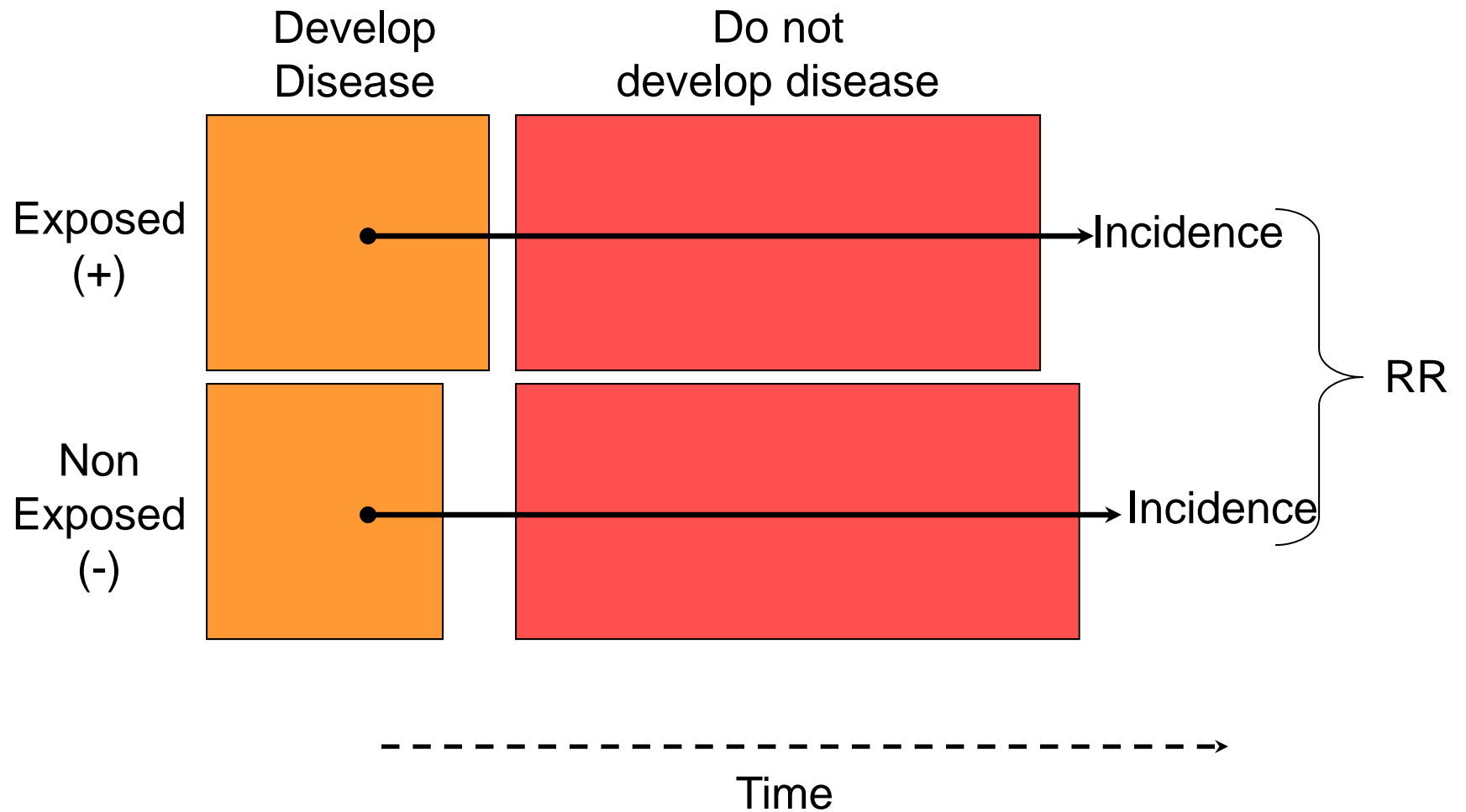




# Design of Cohort Studies



# Cohort studies



# Follow Over Time to Determine Whether

	Disease Develops (+)	Disease does not develop (-)	Incidence rates of disease
<div style="border: 1px solid black; padding: 5px; display: inline-block;">Initially Select Samples</div> Exposed to suspect risk factor(s) (+)	a	b	$\frac{a}{a+b} =$ Incidence in exposed persons
Not exposed to suspect risk factor(s) (-)	c	d	$\frac{c}{c+d} =$ Incidence in non-exposed persons

$$\text{Relative Risk} = \frac{\text{IR of Disease in Exposed Individuals (+)}}{\text{IR of Disease in Non-Exposed Individuals (-)}} = \frac{a/a+b}{c/c+d}$$

# Directions of a Cohort Study

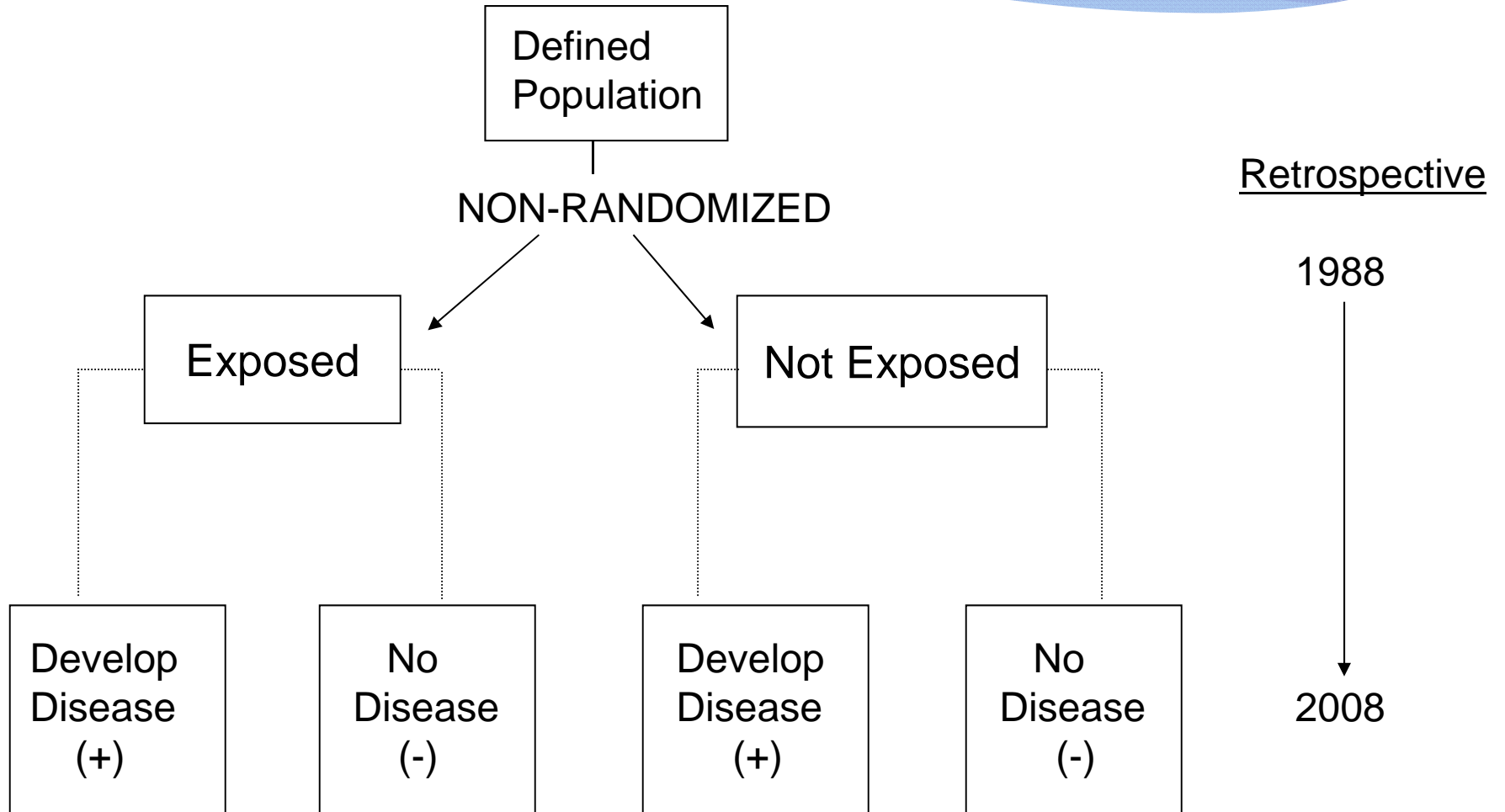
- \* **Retrospective:**

- \* Outcomes of interest and exposures have already occurred when study is initiated

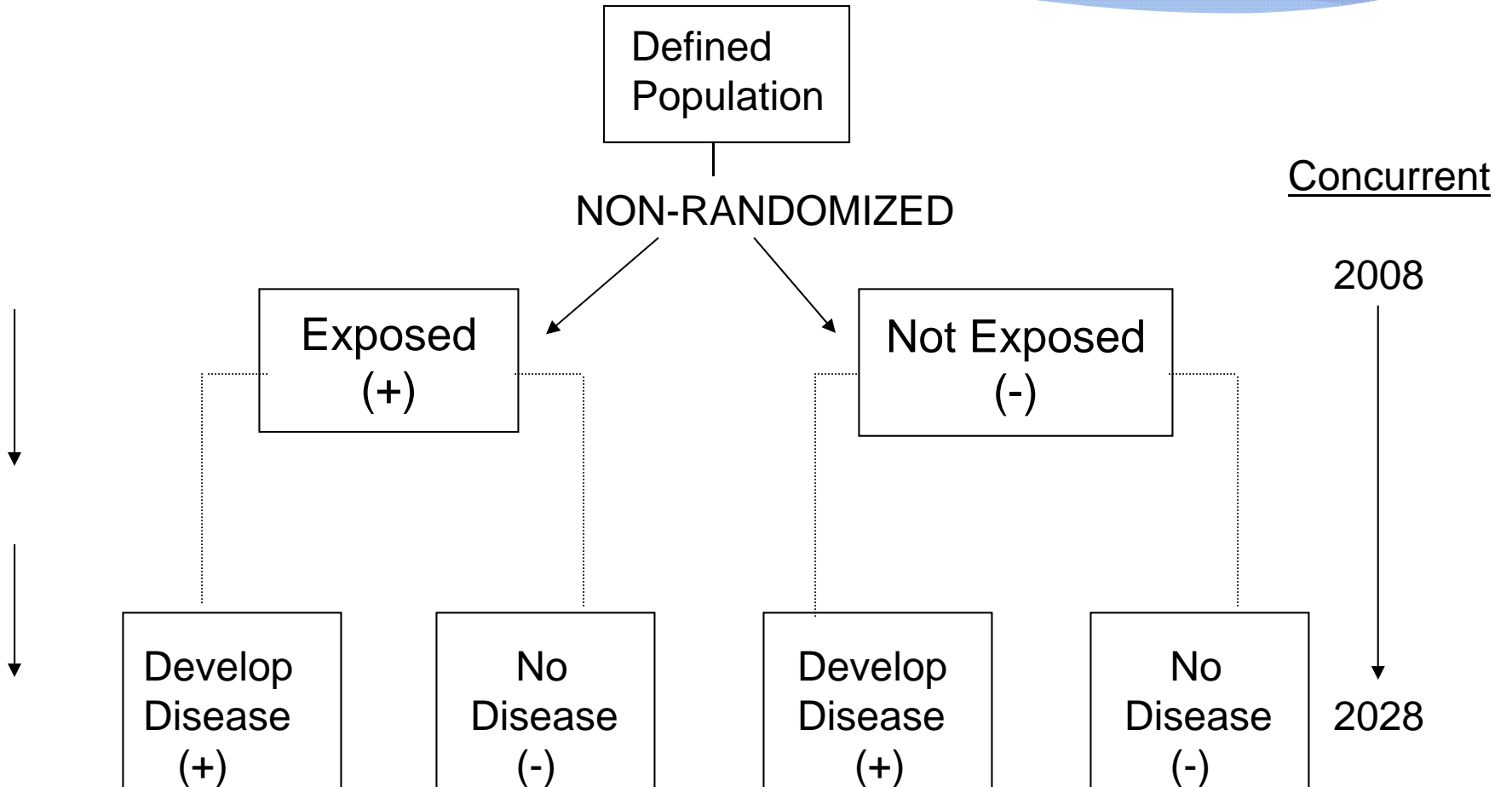
- \* **Prospective:**

- \* Outcomes have not occurred when study is initiated

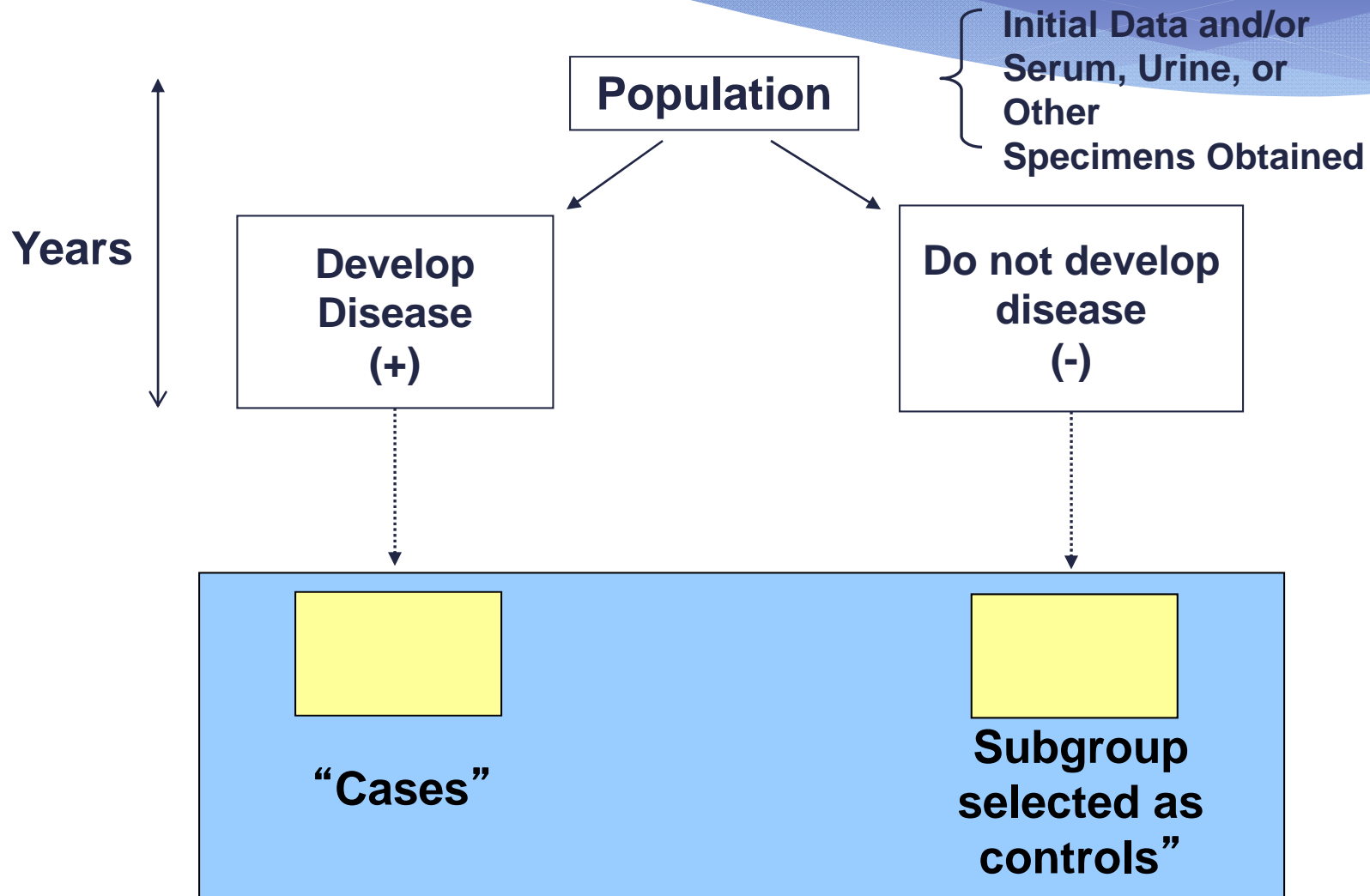
# Time Frame For A Retrospective Cohort Study Begun in 2008



# Time Frame For A Concurrent Cohort Study Begun in 2008



# Nested case-control study



# Framingham Heart Study

- \* Prospective longitudinal cardiovascular study began in 1948 with 5, 209 adult (30-62 yo) study participants.
- \* On it's third generation and has provided the basis for now-common knowledge concerning heart disease such as effects of diet, exercise and aspirin.
- \* Coined the term 'risk factor'.
- \* Lifestyle, environmental factors and inheritance.

**<http://www.framinghamheartstudy.org/index.php>**



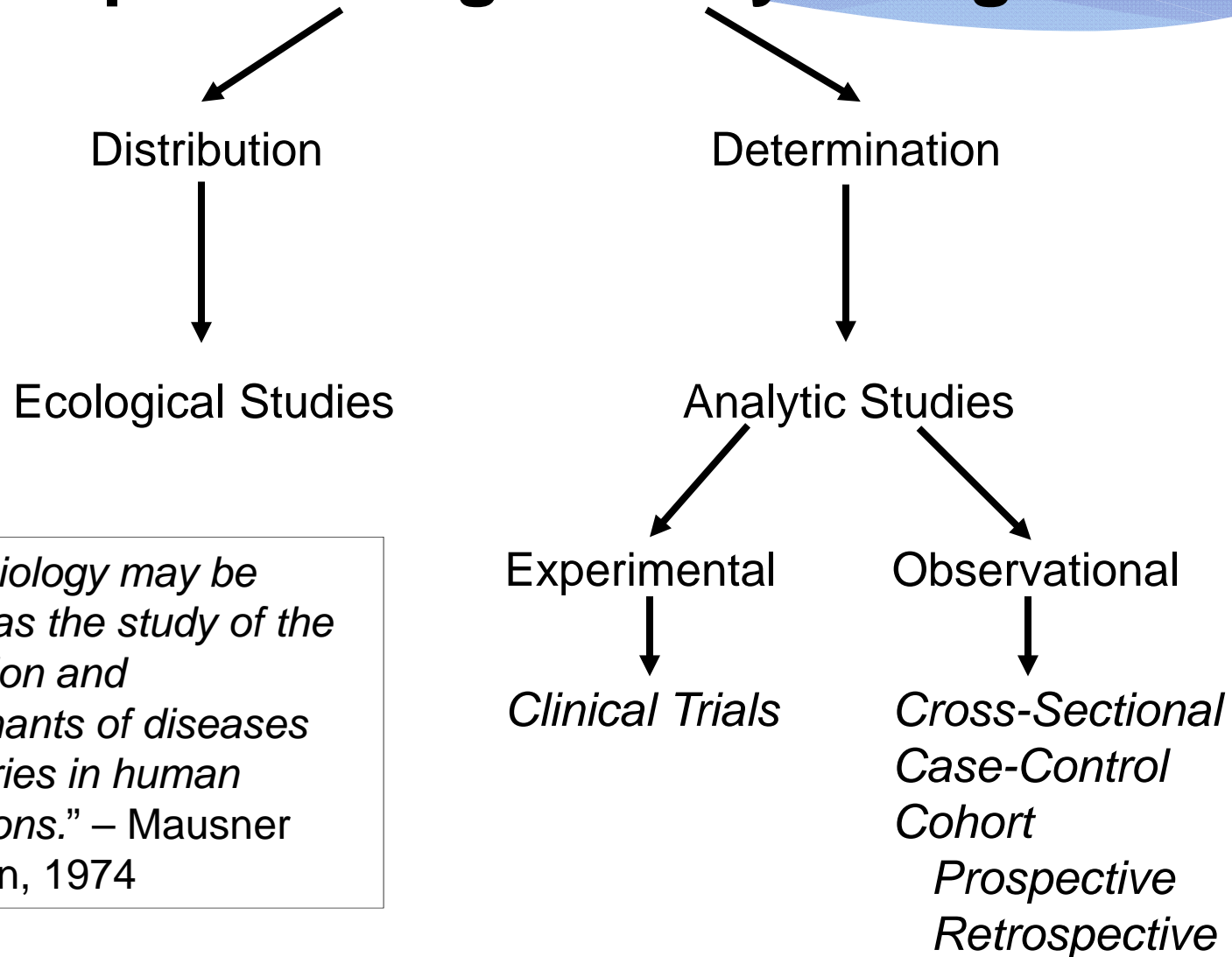
# Advantages of Prospective Cohort Studies

- \* Can directly estimate incidence rate of disease in exposed and non-exposed individuals
- \* Less bias in assessing exposure factor of interest
- \* Good study to do when exposure is rare
- \* Multiple health/disease outcomes can be examined
- \* Temporal relationship between exposure and disease is clear

## Disadvantages of Prospective Cohort Studies

- \* Large study population usually needed
- \* Generally expensive to carry out
- \* Long follow-up is needed in concurrent studies
- \* Potential problem of bias in assessing outcome
- \* Use of retrospective design is possible only if historical data of adequate quality are available

# Epidemiologic Study Designs



*“Epidemiology may be defined as the study of the distribution and determinants of diseases and injuries in human populations.” – Mausner and Bahn, 1974*

# Clinical Trials

- \* Prospective biomedical or behavioral research study of humans to answer a specific question about an intervention (vaccine, treatment, information)
- \* Types of trials: prevention, diagnostic/screening, treatment, quality of life, compassionate use
- \* Designs: randomized, double-blind, placebo-controlled.

# Clinical Trial Phases: each has a purpose and helps scientists answer a different question

**Phase 0** are the first-in-human trials. Single subtherapeutic doses of the study drug are given to a small number of subjects (10 to 15) to gather preliminary data on the agent's **pharmacodynamics** (what the drug does to the body) and **pharmacokinetics** (what the body does to the drugs).

**Phase 1** tests an experimental drug or treatment in a small group of people (20–80) for the first time to evaluate its **safety**, determine a safe dosage range, and identify side effects.

**Phase 2** gives the experimental treatment to a larger group of people (100–300) to see if it is **effective** and to further evaluate its safety.

**Phase 3** gives the treatment to large groups of people (1,000–3,000) to **confirm its effectiveness, monitor side effects**, compare it to commonly used treatments, and collect information that will allow it to be used safely.

**Phase 4** are post-marketing studies to delineate additional information, including the treatment's risks, benefits, and optimal use.

# Clinical Trials

- \* **Strengths**

- \* Directly measures intervention and outcome
- \* Public database of clinical studies

<http://clinicaltrials.gov/>

- \* **Limitations**

- \* Tends to be most expensive type of study
- \* Volunteers may be difficult to recruit – risks have to be minimal or benefit outweighs risks

# Data Collection Basics

- \* Developing survey/data collection instruments does not include just the design of the tool itself.
- \* **Preparing** for the sampling of subjects, collection of data, its entry, editing, storage, and analysis all need to be considered in developing your data collection instruments.
- \* Some of these steps are often overlooked or not given enough attention in the planning of a research study.
- \* There are a number of quality control steps that need to occur both *before* and *during* your data collection instrument development.

# Data Collection Basics: Quality Control Activities

- \* Identify 'needed' data elements
- \* Design forms for recording measurements
- \* Format data collection instruments
- \* Develop coding rules
- \* Develop system for data entry and editing
- \* Develop data analysis plan – it's never too early!
- \* Carry out periodic frequency distributions to assess possible data problems



# Data Collection Basics: Quality Control Activities

- \* Quality of results depends on the **quality of data** collected.
- \* Even the best and thoughtfully developed plans can work out differently once a study is in the field.
- \* A system is needed to maximize completeness and quality of data including pretesting the study steps and/or a pilot study.
- \* Every stage of the research study is a potential source of error – including the development of your survey tool (i.e., thinking about its impact on the validity and reliability of the data you collect).
- \* Anticipate these sources of error and take precautions to minimize them.

# Data Collection Basics: Types of Instruments

- \* Self-administered (e.g., mail, waiting room, internet-based, interactive voice response, etc.)
- \* Telephone (cell phone)
- \* Document review (e.g., medical record abstraction)
- \* Face-to-face interview (translations, back-translations)
- \* Observation
- \* Qualitative studies using focus groups and/or key informant interviews

# Data Collection Basics: Types of Instruments

- \* The particular outcome of interest, the scale of the study, the unit of inquiry, the availability of resources (such as time, money and personnel), and practical considerations are all factors that help determine/guide the type of data collection methodology selected.
- \* Some designs are more prohibitively expensive and/or logistically difficult in certain settings than others.
- \* There are many trade-offs afforded by many available data collection methods: cost-effectiveness, privacy, response rates vs standardization of questions, control over completeness, question order, quantitative vs qualitative information.

# Data Collection Basics: Types of Questions

- \* Behavior (e.g., utilization data)
- \* Attitudes (e.g., opinions, scale scores)
- \* Personal (e.g., demographic data, medical record data)
- \* Environmental (e.g., office practice, seasonal data, geographic data)
- \* Knowledge (e.g., assessment of factual information)

# Data Collection Basics: Types of Question Designs

- \* Different types of questions (e.g., attitudes, beliefs, behaviors, personal data, knowledge-based questions, etc.) may require different design consideration:
  - \* Open-ended questions
  - \* Closed-ended (precoded) questions
  - \* Likert type scales (more likely to less likely range)
  - \* Skip patterns (if yes, go to Section B,)

# Data Collection Basics: Types of Question Designs

- \* All questions in a data collection instrument should be **practical** – ask “why” you need the answer to each and every question; otherwise, it’s easy for a survey or questionnaire to become too lengthy.
- \* Consider both the ability and willingness of study subjects to provide responses – one does not guarantee the other; needs of the respondents may be very different from the needs of the investigator.
- \* Limit the use of vague words (e.g., normally, regularly, generally, usually, on average, overall, etc.).

# Data Collection Basics: One of Many Steps...

Designing a data collection tool is but one of the many important steps in conducting a research study – no matter what setting you're working in.

- \* Defining the study objectives / rationale / research hypotheses
- \* Choosing the study design
- \* Selecting the sample
- \* **Constructing (and pretesting) the data collection tool**
- \* Interviewing the study subjects / respondents
- \* Coding the interviews / forms and entering the data
- \* Analyzing the data
- \* Writing the study report(s)

# Data collection tools

- \* Paper and pen (medical record extraction)
- \* Cell phones and tablets
- \* Electronic medical record systems
- \* REDCap: Research Electronic Data Capture (<http://project-redcap.org/>)
- \* Excel, Access, Open Office
- \* Survey monkey



# Data Analysis Basics:

## Types of Data

- \* **Categorical variables** (also known as “discrete” variables). Categorical variables may be nominal (e.g., gender, race, marital status), ordinal (e.g., education levels, rank orders), or interval (e.g., age groups in 10-year intervals or income in \$10k increments).
- \* **Continuous variables** (also known as “numeric” variables). Examples of continuous variables include: age, years of education, blood pressure, cholesterol level, length of hospital stay, number of ER visits, pharmacy and other ancillary services costs, etc.

# Data Analysis Basics: Managing Data

- \* Data entry
- \* Missing data
- \* Verification of data; i.e., quality control
- \* **Data screening**; i.e., carefully reviewing the data to ensure that they were entered correctly and are being read correctly by the computer. Before conducting any of the more sophisticated analyses, you should carefully screen your data to make sure that you are not analyzing “garbage” (i.e., numbers that were accidentally mis-entered, impossible values on variables that no one could have obtained, and so on). The process of data screening does not guarantee that your data are correct, but it does increase the likelihood.

# Data Analysis Basics: Descriptive Statistics

- \* **Explore the shape of your data.** Among other things, understanding the shape of your data will help you choose the appropriate measure of central tendency (i.e., the mean versus the median). In addition, some statistical procedures require that the sample data be drawn from a normally distributed population, or at least that the sample data do not display a marked departure from normality. You can use the procedures discussed here to produce graphic plots of the data as well as test the null hypothesis that the data are drawn from a normal population.
- \* Provide information that can assist in **decision making, making comparative conclusions, and reporting.**

# Data Analysis Basics: Descriptive Statistics

- \* **Begin to describe data that address your research question(s).** In almost any research article, it is desirable (at a minimum) to report demographic information about the sample studied.
- \* **Descriptive statistics:** means, standard deviations, medians, frequency distributions, percentiles, etc. are used numerically to describe data. Histograms and stemplots may be used to help view the data more graphically. All statistical applications (and even Microsoft Office products: Excel) have procedures to compute descriptive statistics as well as aid in data screening / data management processing.

# Data Analysis Basics: Bivariate Statistics

- \* Bivariate statistics involves the relationship between just two variables; for example, conducting an investigation to study the relationship between obtaining a follow-up mammogram and: race, marital status and age group of women. Here, you would have 3 different bivariate relationships being assessed.
- \* There are a large number of statistical procedures that you can use (in any number of different statistical applications) to investigate bivariate relationships.
- \* Common bivariate statistics include: chi-square tests, t-tests, correlation coefficients, and analyses of variance.

# Data Analysis Basics: Bivariate Statistics

- \* The appropriate bivariate statistic to use depends on the nature of the two variables being studied (i.e., categorical vs continuous) – you need to pay attention to the level of measurement of each variable.
- \* Once you have identified the level of measurement, it's 'relatively' simple to determine the correct statistic for analyzing the relationship between the variables.
- \* Some variables may be measured in various ways (e.g., age and age group); thus, there may actually be more than one statistic that can be used to investigate the relationship between those variables.

# Data Analysis Basics: Bivariate Statistics

- \* chi-square test – 2 categorical variables (disease yes/no vs gender)
- \* t test – 1 continuous and 1 categorical variable with only 2 categories (e.g., blood pressure by race: white vs non-white; ER visits by intervention vs control group)
- \* F test (ANOVA) – 1 continuous and 1 categorical (2 or more categories) variable (e.g., blood pressure and 3+ racial groupings; ER visits by primary care vs staff model HMO vs CHC patients)
- \* correlation analysis – 2 continuous variables (e.g., age and blood pressure; ER visits by years of education or household income)



# Data Analysis Basics: Multivariate Statistics

- \* Multivariate statistics include any of several methods of examining multiple (3 or more) variables at the same time.
- \* These statistics typically have one dependent (i.e., outcome) variable and several independent (i.e., risk factor characteristic) variables.
- \* Whichever statistic you choose, multivariate analyses allow you to examine the relationship between two variables while simultaneously controlling for how each of these may be influenced by other variables.



# Data Analysis Basics: Multivariate Statistics

Some of the more common multivariable / multivariate statistics used in research include:

- \* Logistic regression
- \* Linear regression
- \* Cox proportional hazard models
- \* Survival analyses
- \* Factor analyses

# When to seek IRB approval

Purpose: to assure and protect the rights and welfare of human study participants.

Informed consent and assent (**culturally appropriate**)

Annual review and approval: Tuskegee Syphilis Study

Title 45 Code of Federal Regulations Part 46

Belmont Report and the “Common Rule”

Helsinki Declaration (CITI training online)

Federal Wide Assurance (FWQ) registration with DHHS

Exemptions can be granted only for certain types of studies (check with your IRB for this determination)

International Ethical Guidelines for Biomedical Research Involving Human Subjects: <http://www.hhs.gov/ohrp/international>

# IRB Issues in Global Settings

Research in populations and communities with limited resources:

“Before undertaking research in a population or community with limited resources, the sponsor and the investigator must make every effort to ensure that: 1) the research is responsive to the health needs and the priorities of the population or community in which it is to be carried out; and 2) any intervention or product developed, or knowledge generated, will be made reasonably available for the benefit of that population or community.”

# IRB Issues in Global Settings

Specifics of conducting studies which should be taken into consideration (per UMMS IRB office):

- \* If 'investigator' is doing work as a 'representative' of UMMS, the medical school's IRB needs to be involved; i.e., they have oversight of the decision regarding what level of IRB review is needed.
- \* Also needed is local oversight (i.e., permission from the local clinic/institution – whether a letter of support, a review document from an ethics board, etc. – acknowledging the study and any informed consent issues).
- \* UMMS IRB might 'exempt' it from review, but local oversight in the global setting is still needed.

# IRB Issues in Global Settings

Specifics of conducting studies which should be taken into consideration (per UMMS IRB office):

- \* Fact sheets and written consent forms – translated into a language understandable to the local community (including the appropriate reading level).
- \* UMMS IRB consent template may not be applicable to some global settings (must include the basics of informed consent if not using their template; e.g., purpose, procedures, risks, alternatives, etc.).
- \* Data collection forms – translated (including back translations in some instances); IRB office must have a copy of the English translation for all documents (data collection forms and consent forms).

# Research in Global Health setting

- \* Building human capacity (training)
- \* Building research capacity (technology transfer)
- \* In-country IRB approval (Time!)
- \* Harmonizing clinical practice with research goals in limited resourced settings (translational research)

# Research – how to get started?

- \* Develop your research ideas – it starts with an observation and a problem to solve.
- \* Find a mentor and build your research team
- \* Managing your time (protected time v. delegation)
- \* Get money to fund your project

# Funding research

- \* Depends on disease being studied
- \* Depends on persons being studied
- \* Depends on funders priorities
  
- \* Lists of RFA sent out by University – list can be catered to your area of interest



# UMass Medical School's Office of Global Health

UMass Medical School: Office of Global Health:  
<http://www.umassmed.edu/globalhealth/index.aspx>

Looking for projects, grants and funding:  
<http://www.umassmed.edu/Content.aspx?id=150692>

American Society of Tropical Medicine and Hygiene  
<http://www.astmh.org/Home.htm>

## **UMMS Office of GLOBAL HEALTH GRANTS**

- \* Pilot Project Grant (\$35k)
- \* Travel Awards (\$5k)

# Questions?



Acknowledgments for slides:  
Rob Goldberg, QHS  
Judy Savageau, Family Med

# Questions for discussion

- \* How does research differ from clinical practice?
  - \* **Individual** medical history to prescribe treatment
    - \* Diagnostics and application of known treatments or interventions
  - \* **Population-based** data on groups of people to generate hypotheses about links between exposures and disease
    - \* Epidemiologic study design to discover unknown links
    - \* Translational goals

# Interpreting Study Findings

- \* There are a number of issues involved in interpreting the findings of a single epidemiological study.
- \* Interpreting the findings of a single study includes considering the strength and precision of the effect estimate and the possibility that it may have been affected by various possible biases (confounding, selection bias, information bias).
- \* If it is concluded that the observed associations are likely to be valid (including statistically significant), then attention shifts to more general causal inference, which should be based on all available information.
- \* Epidemiologic studies almost always contain potential biases, and the focus should be on assessing the likely direction and magnitude of the biases, and whether they could explain the observed associations.

# Interpreting Study Findings

- \* If it is concluded that the association in a particular study is unlikely to be primarily due to bias and/or chance, attention then shifts to assessing whether this association exists more generally, and whether the association is likely to be causal.
- \* Sir Bradford Hill's criteria (Hill, 1965) for causation are a group of minimal conditions necessary to provide adequate evidence of a causal relationship between an exposure/risk factor and a disease/illness/outcome.
  - \* Considerations for assessing the epidemiological evidence include: temporality, biologic plausibility, specificity, consistency, strength of association, reversibility, and whether there is evidence of a dose-response relationship.

# Interpreting Study Findings

Despite the continual need to assess possible biases, and to consider possible imperfections in epidemiologic data, it is also important to ensure that preventive action occurs when this is warranted, albeit on the basis of imperfect data. Hill wrote:

"All scientific work is incomplete - whether it be observational or experimental. All scientific work is liable to be upset or modified by advancing knowledge. That does not confer upon us a freedom to ignore the knowledge that we already have, or to postpone the action that it appears to demand at a given time."

# Interpreting Study Findings

- \* Results need to be interpreted in an objective and critical way before assessing their implications and before drawing conclusions.
- \* Interpretation of research results is not just a concern for researchers. Health professionals learning about research results should be able themselves to interpret them correctly, and to assess their implications for their work.
- \* Policymakers should also be aware of the possible pitfalls in interpreting research results and should be cautious in drawing conclusions for policy decisions.



# Interpreting Study Findings

- \* Interpreting results of *qualitative research* involves the interpretation of textual materials taken from talk or observation.
- \* In interpreting qualitative findings, investigators need to carefully look into issues of: credibility, dependability, confirmability, and transferability.
- \* NOTE: Lots of materials/resources can be made available for qualitative research studies – not included in any depth here – *but please ask if it's a particular interest of yours!*



## Checklist of Factors to Consider in Reviewing Design and Analysis of Case-Control Studies

	<u>Yes</u>	<u>No</u>	<u>Uncertain</u>
• Is case-control design appropriate for hypothesis being examined or would another study design be more appropriate or cost-efficient?	✓	<input type="checkbox"/>	<input type="checkbox"/>
• Is sampling method for selection of case and control groups clear and understandable?	✓	<input type="checkbox"/>	<input type="checkbox"/>
• Have case and control groups been selected without regard to exposure factor(s) of interest?	✓	<input type="checkbox"/>	<input type="checkbox"/>
• Are incident cases of disease being studied or are only prevalent cases included in whom the etiology-disease association might be more questionable?	✓	<input type="checkbox"/>	<input type="checkbox"/>

## Checklist of Factors to Consider in Reviewing Design and Analysis of Case-Control Studies

	<u>Yes</u>	<u>No</u>	<u>Uncertain</u>
• Do cases under study reflect a range of disease severity or are only select cases from a single hospital or ambulatory care clinic represented?	✓ <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Were appropriate hospital, clinic, and/or neighborhood controls selected?	✓ <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Have a sufficient number of cases and controls been selected to adequately test the proposed hypotheses?	✓ <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Have cases and controls been matched on a limited number of other relevant factors and have satisfactory matches been achieved?	✓ <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

## Checklist of Factors to Consider in Reviewing Design and Analysis of Case-Control Studies

	<u>Yes</u>	<u>No</u>	<u>Uncertain</u>
• Has categorization of the primary exposure factor(s) been carried out into appropriate dose/response categories?	✓ <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Were individuals involved in either direct data collection or record abstraction blinded to study hypotheses and/or to case/control status?	✓ <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Are OR' s and accompanying confidence intervals presented to quantify these disease: exposure association?	✓ <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Have potential confounding variables been adequately considered and controlled for analytically?	✓ <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

## Factors to Consider in Reviewing Design and Analysis of Longitudinal Studies

	<u>Yes</u>	<u>No</u>	<u>Uncertain</u>
• Is use of a prospective design appropriate for hypothesis being examined?	✓ <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Have exposed and nonexposed study subjects been selected from similar populations? Has investigator presented a sufficient rationale for choice of the study population?	✓ <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Has exposure status been adequately measured and independently validated?	✓ <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Have possible changes in exposure status since time of initial baseline classification been measured and taken into account?	✓ <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

## Factors to Consider in Reviewing Design and Analysis of Longitudinal Studies

	<u>Yes</u>	<u>No</u>	<u>Uncertain</u>
• Have study endpoints been determined without regard to exposure status. Are persons involved in collection of these data unaware of primary study hypotheses?	✓ <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Has determination of principal study outcome been adequately measured and independently validated?	✓ <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Have potentially confounding factors been measured? Has influence of these factors been controlled for analytically?	✓ <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
• Has an acceptable means of determining subject follow-up been used and has a high follow-up rate been achieved?	✓ <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

## Factors to Consider in Reviewing Design and Analysis of Longitudinal Studies

	<u>Yes</u>	<u>No</u>	<u>Uncertain</u>	
<ul style="list-style-type: none"> <li>• If an acceptable follow-up rate of exposed and nonexposed cohorts has not been achieved, have the sociodemographic or clinical characteristics of those unavailable for follow-up been compared with those remaining under follow-up to determine whether groups are comparable?</li> </ul>	✓	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<ul style="list-style-type: none"> <li>• Did study have adequate power to detect differences in principal study outcome(s) in exposed and nonexposed cohorts?</li> </ul>	✓	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<ul style="list-style-type: none"> <li>• Was duration of follow-up sufficient?</li> </ul>	✓	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>