Welcome to Section II

- Amy R. Borenstein, Ph.D., M.P.H.
  Professor of Epidemiology
  - Room 2109 Friday 2-4pm 974-6670
  - aborenst@health.usf.edu
  - Ph.D. from University of Washington, Seattle (1988)
  - M.P.H. from University of Michigan, Ann Arbor (1980)
  - Professional history: 1 year at EPA/Washington D.C., 1 year at Versar, Inc post-Master’s
  - Switched from env. Epi to chronic disease; PhD at UW; worked 9 years at Battelle Memorial Institute; left 1996 for USF

Welcome to Section II

- Research Interests:
  - 24 years in neuroepidemiology
  - Focus on Alzheimer’s disease, vascular dementia, mild cognitive impairment
    - Among first case-control studies in literature (dissertation)
    - Case-control study of the potential role of aluminum in AD
    - Cohort study of 1,985 Japanese-Americans (1991-2002), Kame Project, part of the Ni-Hon-Sea cross-national study of dementia and its subtypes among individuals of Japanese descent
    - Cross-sectional studies of African Americans (Hillsborough Elder African American Life Study) and (prospective) Caucasians (Charlotte County Healthy Aging Study)
    - Currently working on collaborations to set up a large cohort study of about 10,000 community-dwelling residents in Shanghai, China to study risk and preventive factors for brain atrophy, mild cognitive impairment and dementia
    - Interests in gene-environment interactions, early detection of dementing illnesses and finding at-risk subjects in asymptomatic stages; ethnic differences in etiologic associations for dementia
  - Multiple sclerosis
  - Inflammatory Bowel Diseases (IBD)

Study designs

- Observational
  - Descriptive:
    - Case report
    - Case series
    - Cross-sectional/Correlational/ecologic
  - Analytic:
    - Prospective cohort
    - Retrospective (or historical) cohort
    - Case-control
    - Nested case-control

- Experimental
  - Clinical trials
  - Intervention trials
  - Prevention trials

Measures of association

- Analytic design
  - “a study designed to examine associations, commonly putative or hypothesized causal relationships” Last JM, ed. A Dictionary of Epidemiology, 3rd edition. NY, Oxford University Press, IEA, Inc, 1995

- Does Exposure (E) cause disease outcome (D)?
  - \[ P(y) = x \]
  - \[ P(D) = E \]

Associations

- The extent to which things tend to occur together (non-directional)
  - OR
  - The statistical dependence between two variables
  - In epidemiology, we would like to establish causal associations (uni-directional)
Associations, con’d

- Prevalence ratio or odds ratio can tell us what the association is between E and D but can not necessarily tell us which came first (E→D or D→E, “chicken-or-egg”)
- Under some circumstances, e.g., when E is a gene, we know that E precedes D in time (except when D is congenital!)
- We only use a case-control approach when cohort studies are prohibitively expensive (rare disease) and/or when we have an expensive assay for exposure (nested case-control)

Philosophical (conceptual) point: in a sense, all epi is based on the cohort study design

Under some circumstances e.g., when E is a gene, we know that E precedes D in time (except when D is congenital!)

We only use a case-control approach when cohort studies are prohibitively expensive (rare disease) and/or when we have an expensive assay for exposure (nested case-control)

“The basic building blocks for epidemiologic inference are incidence rates”.
Ken Rothman and Sander Greenland, Modern Epidemiology, 2nd Ed, 1998

Other factors

- We also must understand how other variables (z, ≠) are related (either non-directionally or directionally) to x and y in order to understand the main association between x and y.
  (confounding, effect-modification) Unit 8

Causal pathways

- Exposure (E) → Disease (D)
- Other factor 1
- Other factor 2
- Other factor n

Definition of a “cause”…

- “Characteristic, event or condition that preceded the event and without which the disease event would not have occurred at all or would not have occurred until some later time” (Rothman & Greenland, ME II 1998)
- Temporal sequence is the key.

Web of Causation Leading to Coronary Heart Disease

http://classes.iuuc.edu/sonmed902/aging_principles/demography/rectangularization_of_mortality-popup.htm
Essential concept behind the prospective design (cohort, historical, randomized controlled trial):

- Exposures are identified first
- Population is followed over time
- Outcomes/disease(s) are assessed over time
- Incidence rates are compared between E and $E$
groups
- Experimental study: E is manipulated by investigator

$$\text{Exposure} + E \quad \text{Time} \quad \text{Rate of disease}$$

$$\text{Exposure} - E \quad \text{Time} \quad \text{Rate of disease}$$

We would like to emulate an experimental study design…

- Experimental animal studies:
  - Breed mice so genes are controlled. Control environment in which mice live.
  - Have enough mice so don’t make the mistake of not having sufficient sample size to find something if it exists.
  - Randomize mice to treatment (E) and placebo groups ($E$)
  - Administer controlled doses of medication by mouth (diet)/injection/topically (compliance is assured)
  - Compare clinical disease in Rx and $Rx$ mice
  - Sacrifice animals, look for pathologic disease in Rx and $Rx$ mice

Three kinds of experimental studies in epi:

- Randomized controlled trials (single- or double-masked)
  - Usually conducted among patients to test drug efficacy,
  - or, among high-risk individuals (e.g., MCI to AD) – secondary/tertiary prevention
- Intervention/prevention trials
  - Conducted among asymptomatic individuals (much larger numbers needed) – primary/secondary prevention
- Community-based trials
  - Unit of measurement is not the individual but a whole community (e.g., fluoride), neighborhoods, schools (classrooms receiving intervention) - primary/secondary prevention

Experimental studies vs. observational studies: In people, it would be unethical to randomize most E or Rx

- Main difference is that E is assigned randomly
- Elements of RCTs that are under investigators’ control:
  - Size of sample
  - Proper randomization
  - Some measure of compliance during follow-up
  - Design trial with long enough follow-up period that would result in observable difference in incidence of outcome between two (or more) groups
Experimental studies vs. observational studies

- Elements of RCTs that are not under investigators’ control:
  - People may not comply with treatment
  - People drop out of the study: death, refusal, move out of study area
  - People are inherently heterogeneous, unlike bred animals
  - People don’t live in controlled environments like experimental animals
  - There may be variations by ethnicity so RCT may only apply to certain subpopulations studied

When E or Rx cannot be randomly assigned, a prospective observational study is the next best thing.

- People “choose” to be exposed (behavior, work, diet) or are exposed accidentally (2nd hand smoke, chemical exposures)
- We can follow people over time to see if those who are exposed are more or less likely to develop a certain outcome(s).

Problem:
- People who are exposed may be different than those not exposed.
- We adjust for the differences between these two groups but sometimes we cannot adjust for all of the factors, or the factors may be unmeasurable (intrinsic confounding)

Sometimes a case-control study is better than a cohort study!

- In 1950’s and 1960’s, diethyl stilbestrol (DES), a hormone used to prevent miscarriages, was given to women during pregnancy
- Case report in 1970 of 7 cases of vaginal clear cell adenocarcinoma (CCAC) in the daughters of mothers who took DES
- Case-control study in 1971 (Herbst, Ulfelder, Poskanzer NEJM 284:878-81) of 8 cases of CCAC and 4 matched controls (32 controls matched on birth within 5 days of proposita in same hospital and same type of service (SES)):

<table>
<thead>
<tr>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Est +</td>
<td>7</td>
</tr>
<tr>
<td>Est -</td>
<td>1</td>
</tr>
</tbody>
</table>

Example

- Prospective study attempted at Mayo Clinic in which about 900 exposed girls and 800 unexposed yielded no cases in the exposed (insufficient sample size to find even 1 case given CIR=1/1000/yr).
- Turns out incidence rate (per year) of CCAC is 1/1000 among DES-exposed fetuses.
- Had a prospective study been conducted first, this association would not have been discovered

Example

- Disease-free individuals are selected and their exposure status is ascertained
- Subjects are followed for a period of (real) time to record the development of new disease occurrence (incident cases)
Historical cohort study (retrospective cohort)
- Exposure status (data) is documented for the beginning of the observation period in a disease-free cohort
- Incidence of disease is ascertained over the follow-up period.
- Both exposure and disease have already occurred, but the sample is selected by exposure status (not real time)

Example of a prospective cohort vs. historical cohort study
1. Prospective: investigator accompanies Ss through time
   - 2007: Identify elementary school students (measure baseline characteristics @ ages 5-10)
   - 2017-2022: Measure smoking status at ages 15-20
   - Measure lung diseases at ages 55-60
   - Investigator is dead

2. Historical cohort: investigator can do entire study in 1 year
   - 1942: Identify elementary school students
   - 1960: Survey of smoking habits
   - 2007: Traces outcomes using Medicare files
   - Investigator is still young and has a long career ahead of him

Importance of removing prevalent cases to have a disease-free cohort
- Baseline ($T_0$)
- Wave 1, n ($T_n$)
- Remove prevalent cases (P=Incidence X Duration)
- Enumerate incident cases

Ambi-directional cohort study
- Both prospective and retrospective components:
- Incidence of disease is ascertained over the follow-up period.
- We can monitor exposure multiple times during the follow-up (all types of cohort studies)
- Disease is monitored both before (not real time) and after (in real time) study is initiated – sometimes less accurately before: e.g., HHP vs. HAAS
- Sample is selected by exposure status

Every prospective study becomes a historical cohort study if you wait long enough...
- When data collection is over and the study is terminated (usually because funding has ceased), a prospective cohort study becomes a retrospective cohort study, i.e., all the data have been collected.

Choose a cohort design when:
- The outcome is relatively common
- The exposure is relatively rare
- You want to study multiple outcomes of an exposure
- You are able (or try hard) to minimize loss to follow-up
- The interval between E and D is relatively short or when you think you can follow the cohort for a long enough time to account for the latent period (induction period) of the disease
- You want to be confident about temporal sequence (that D follows E in time)
- You want to study a particular cohort (e.g., The Nun Study, The Kame Project)
Participants: School Sisters of Notre Dame
- Nuns living in U.S. who were born <1917 were invited to participate in 1991-1993 in a prospective study of mental health and physical function
- Inclusion criteria: agree to be evaluated annually, provide access to archival and medical records and to donate brain upon death for neuropathologic studies
- Of 1,030 Sisters eligible and invited, 678 (66%) agreed to participate
- Nuns were age 75-102 at baseline

**The Nun Study**

The Gores Girls in midlife:
In some cases our sisters are also siblings.

![A normal brain](image)

![An Alzheimer brain](image)

**The Ni-Hon-Sea Project: study hypothesis**

<table>
<thead>
<tr>
<th></th>
<th>AD pathalogy</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Demented during life</td>
<td>223</td>
<td></td>
</tr>
<tr>
<td></td>
<td>81</td>
<td>374</td>
</tr>
<tr>
<td></td>
<td>407</td>
<td>678</td>
</tr>
<tr>
<td>RR(95) =0.852/0.1780</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.79</td>
</tr>
</tbody>
</table>

Japan | Hawaii | Seattle

**Sources of identification of 3,196 persons aged 65+ in Kame census, 1991**

<table>
<thead>
<tr>
<th>Source</th>
<th>n eligible (%)</th>
<th>n completing baseline exam (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Telephone directory</td>
<td>2,084 (65.2)</td>
<td>1,375 (66)</td>
</tr>
<tr>
<td>Voluntary</td>
<td>82 (2.6)</td>
<td>70 (85)</td>
</tr>
<tr>
<td>Japanese-American</td>
<td>186 (6.8)</td>
<td>127 (68)</td>
</tr>
<tr>
<td>Citizen's League</td>
<td>87 (2.7)</td>
<td>56 (64)</td>
</tr>
<tr>
<td>Keiro nursing home</td>
<td>193 (6.0)</td>
<td>142 (74)</td>
</tr>
<tr>
<td>Nikkei Concerns</td>
<td>76 (0.5)</td>
<td>51 (67)</td>
</tr>
<tr>
<td>Referred by friend/relative</td>
<td>8 (0.3)</td>
<td>6 (75)</td>
</tr>
<tr>
<td>Family history quest.</td>
<td>122 (3.8)</td>
<td>42 (34)</td>
</tr>
<tr>
<td>Obituaries</td>
<td>62 (1.9)</td>
<td>0 (---)</td>
</tr>
<tr>
<td>Hiroshima Bomb Study</td>
<td>1 (0.03)</td>
<td>1 (100)</td>
</tr>
<tr>
<td>Family identified from obits</td>
<td>3 (0.09)</td>
<td>2 (67)</td>
</tr>
<tr>
<td>HCFA Medicare ben</td>
<td>292 (9.1)</td>
<td>121 (41)</td>
</tr>
<tr>
<td>Total</td>
<td>3,196</td>
<td>1,993 (62)</td>
</tr>
</tbody>
</table>

**Kame Project - Summary of Methods: prevalence and incidence phases**

Nov 1, 1991: Census of all Japanese Americans aged 65+ in King County, WA
N = 3,196

First Biennial 1994-1996 n = 1,836

Biennial 2,3,4 1996-2001

Refused n=979
Died n=  61
Unable to reach, moved n= 20

Participants in baseline exam
Phase I: 1992-1994 N = 1,985

Sampled by age and CASI score

Clinical Examinations n=382

Score < 877

Non-sampled and non-demented from clinical exam

Annual track n=149

Clinical Examination
Translation…

At end of study period

- 149 prevalent cases of dementia found at baseline
- 172 incident cases of dementia found over eight-year follow-up period among non-demented cohort
- Over 10,000 person-years accumulated
- Analyses examine how risk factors from baseline predict incident cases (best to study etiology since $P = I \times D$).
  - Prevalence does not measure risk.

Articulate Quizmaker Quiz Placeholder - review2_lecture5_part1

Calculation of the RR

- Cumulative incidence rate
  \[
  \text{Cumulative incidence rate} = \frac{\# \text{ new cases}}{\text{population at risk}} \times \text{over a specified time period (usually 1 year)}
  \]

  *if # new cases/population at risk occurs over several years, need to divide by # years to get annual rate

- Incidence density rate
  \[
  \text{Incidence density rate} = \frac{\# \text{ new cases}}{\text{person-years/time at risk}}
  \]

Calculate the RR

<table>
<thead>
<tr>
<th></th>
<th>$D$</th>
<th>$\bar{D}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E$</td>
<td>54</td>
<td>432</td>
</tr>
<tr>
<td>$\bar{E}$</td>
<td>75</td>
<td>1,414</td>
</tr>
</tbody>
</table>

\[
\text{RR} = \frac{54/486}{75/1,489} = \frac{0.111}{0.0504} = 2.21
\]

$4860 \text{ py}$

<table>
<thead>
<tr>
<th></th>
<th>$D$</th>
<th>$\bar{D}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E$</td>
<td>54</td>
<td>4860 py</td>
</tr>
<tr>
<td>$\bar{E}$</td>
<td>75</td>
<td>14,890 py</td>
</tr>
</tbody>
</table>

\[
\text{RR} = \frac{54/4860}{75/14,890} = \frac{0.0111/0.00504}{2.21}
\]

Do not learn Oleckno’s way of computing person-time (pp. 69-70)

- Calculate denominator (total number of individuals (CI) or person-years of health (ID))
- Calculate numerator (number of events)
- Express per 1,000, 10,000, etc.
- Remember time unit for incidence and prevalence rates.
Person-time
- Person-time accumulates when we observe a group of individuals over a period of time to ascertain the development of an event
- Although we would like to follow everyone in the sample indefinitely, (cumulative incidence), the actual time each individual is observed will most likely vary, since:
  - Subjects may be recruited at different times
  - Subjects migrate (move)
  - Subjects choose to leave study (dropout)
  - Subjects die
  - Subjects develop the disease of interest
  - The study ends

Compare two incidence rates:
- Stratify by exposure
  \[
  \frac{\text{Incidence rate in exposed}}{\text{Incidence rate in unexposed}} = \text{RR}
  \]
- RR = 1.0  Null value
- RR > 1.0  Increased risk (risk factor)
- RR < 1.0  Decreased risk (protective; inverse)

RR, HR, OR are all on log scale (ln):

OR vs. RR
- Do not usually calculate ORs in cohort studies because we have rates. OR is usually used to approximate the RR
- If we have rates, we use them.

Please continue to...
Cohort Studies - Part 2