Epidemiology 2200b
Class #3

Case-control, cohort and cross-sectional studies

Reading: Gordis Ch. 9, 10 &13

Review of previous lecture

1. States and Events
2. prevalence and incidence,
   Mortality and recovery
3. rates and risk
4. relative rates and risk
5. Population and samples

Objectives of *Analytic* Epidemiological Research

- To provide estimates of the association between Exposures (E) and Outcomes (O)
  - risk factors for for disease
  - protective factors and effective Treatments
- To judge whether an association is
  - result of confounding
    - caused by something other than E
  - result of chance alone
    - no causal relationship
  - possibly causal (within range of certainty)

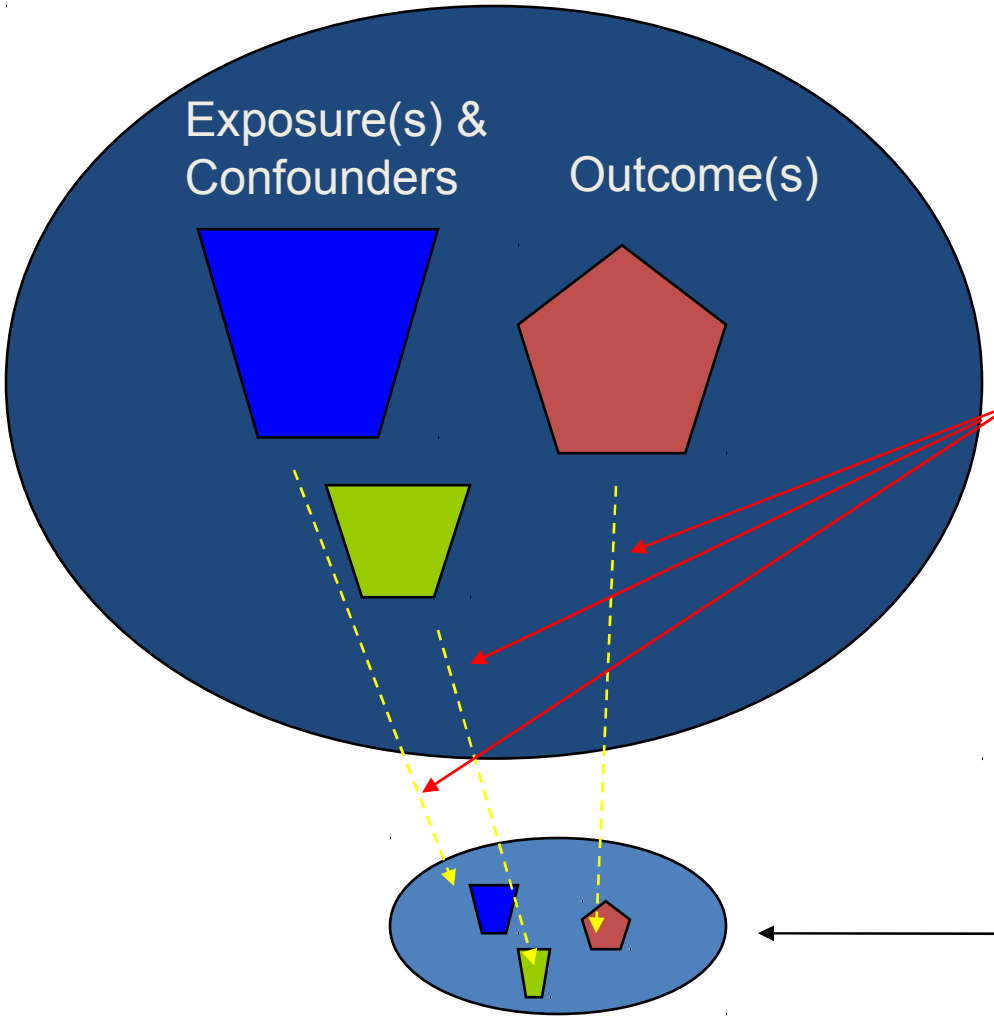
Today's Objectives

- Define basic division between *observational* and *experimental* designs
- Identify *case-control, prospective and retrospective cohort designs* from study descriptions
- List main strengths and limitations of each
- List methods to control for imbalances between groups
- Define *retrospective* vs. *prospective* designs and measures
- Examine how the OR works (*observed* vs*. expected*)
- Calculate Risk Ratio for 'dose response'
- Critically appraise some real examples

Controlled Observations are Fundamental to Science

- Two basic choices in epidemiology:
- Control on basis of outcome
  - Controlled observations between diseased (cases) and well (controls)
- Control on basis of exposure
  - Controlled observations between exposed and unexposed
- Ideal: groups identical on all factors other than factors of primary interest

The Theoretical Ideal Underlying Observational Epidemiology Studies



Population of Interest

Unbiased sampling & measurement

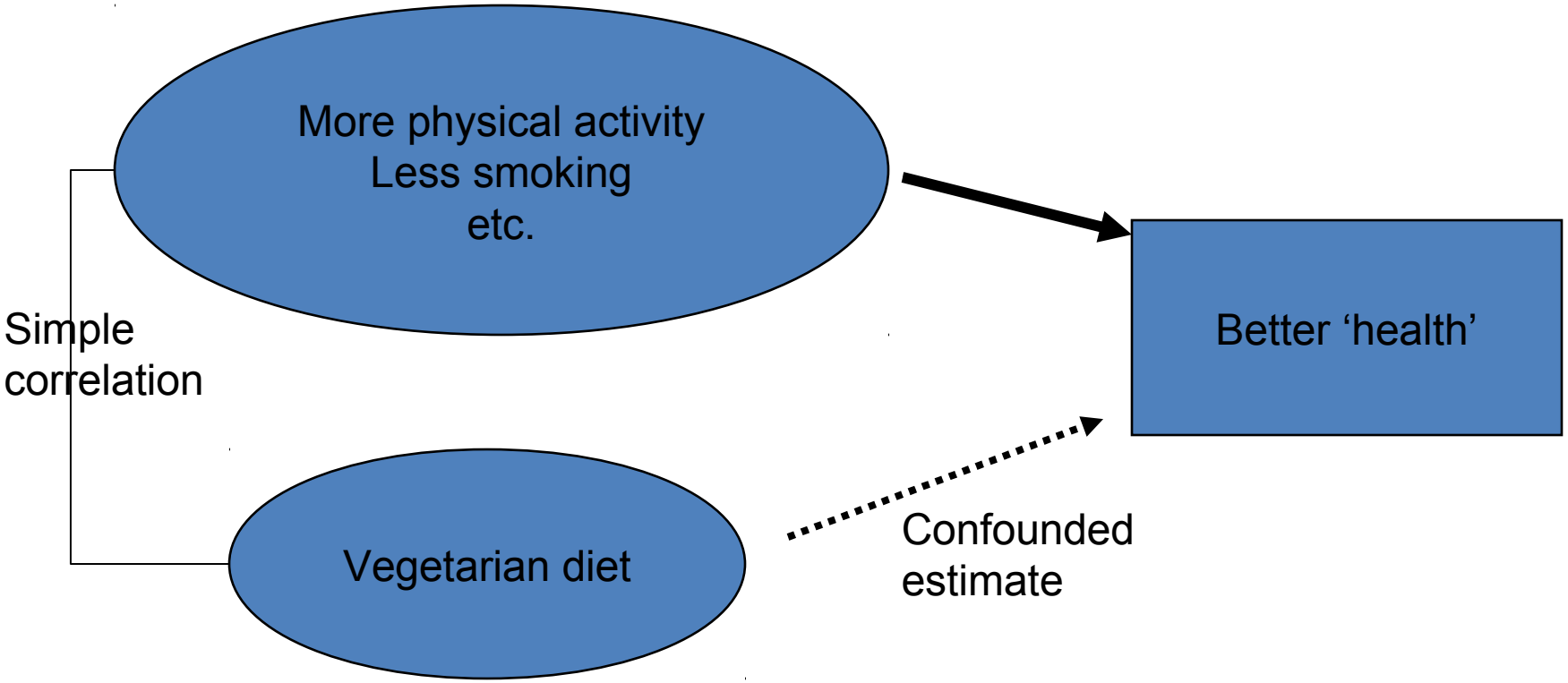
Yields…

Study Sample

Representative exposures and outcomes

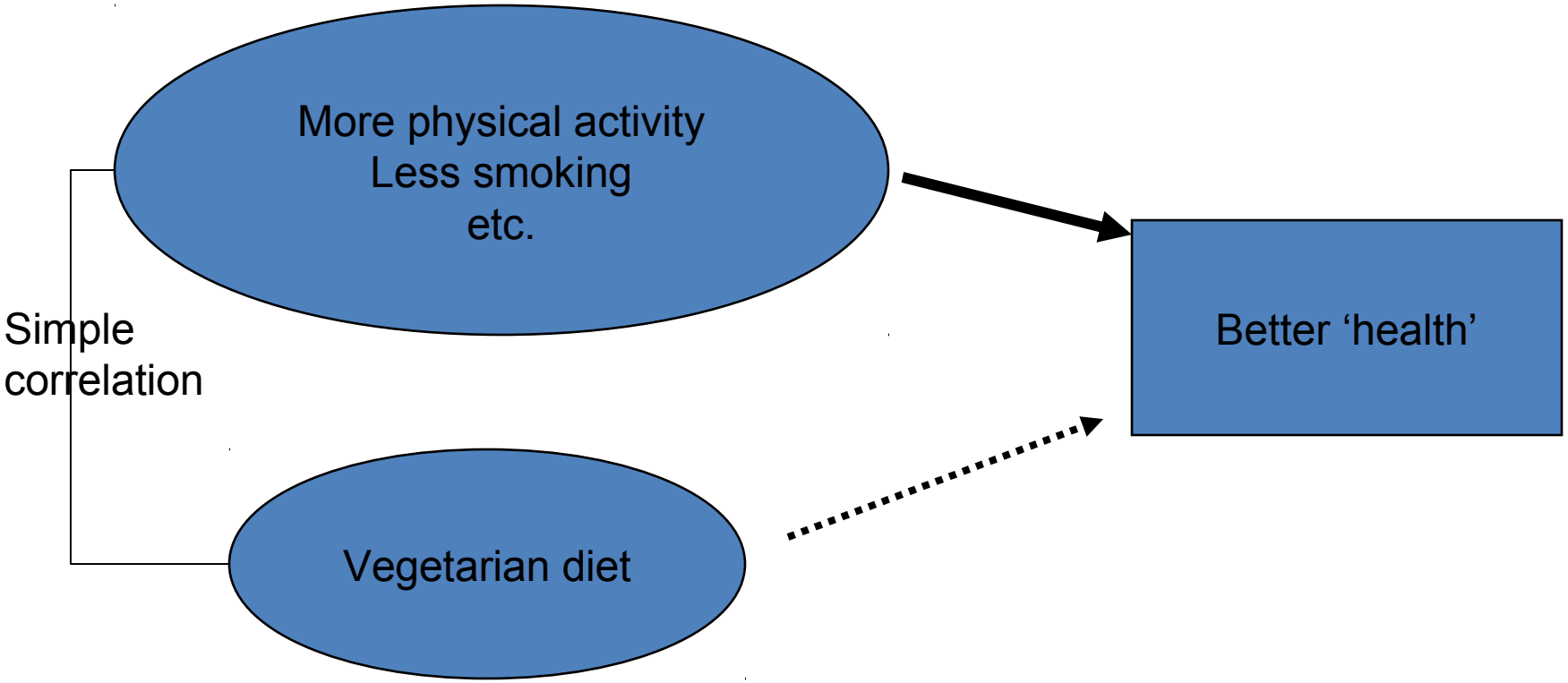
the study sample is representative of the population then the estimates should be represe

Confounding Example



__If__ vegetarians more likely to exercise, less likely to smoke (and __if__ these are true causes of better 'health'), the crude protective effect of a vegetarian diet on better 'health' is said to be confounded by these health practices

Correlation = imbalanced groups



'Correlation' between diet and health practices means diet groups are imbalanced on health practices: e.g.
Non-vegetarians, 40% smoke; Vegetarians, 10% smoke

Controlling for Imbalances Between Groups[1]

Study design stage:

- Randomization
  - random assignment or random allocation
  - to veg/non veg (% of smokers distributed by chance)
- Matching
  - match on smoking status (Y/N) in veg/non veg groups
- Restriction
  - study only non-smokers (veg/non veg)

Analysis stage:

- Stratification
  - Stratify by smoking status and calculate Odds/Risk Ratios of veg/non-veg in each stratum of smoking status
- Multivariable statistical models
  - logistic or linear regression that adjust for imbalances among several confounders

[1]Examples assume imbalance in smoking status between vegetarians/other diets

Failure to Control for Confounding

Rothman, Epidemiology, 2002:3

Risk of death in 20 years of follow-up among women in Whickham, England, by baseline smoking status (n = 1314)

Smokers: 139 deaths/582 women Risk = 0.24
Non-Smokers: 230 deaths/732 women Risk = 0.31

Crude Risk Ratio (95%CI)= 0.24/0.31 = 0.76 (0.63- 0.91)

Smokers are 24% *less* likely to die over 20 years of follow-up (p = 0.002)!!!

What **single** factor would you suspect first as a possible confounder of this relationship?(Formal solution in Class 6)

Examples of Research Questions

- Does ginseng consumption reduce cancer risk?
- Does exposure to agricultural pesticides increase risk of fetal deaths due to congenital anomalies?
- Does Highly Active AntiRetroviral Therapy (HAART) lengthen the disease-free time in HIV infection?

For each of the above:

1. Is there an effect? (non-null OR/RR)
2. In which direction? (> 1.0 risk or < 1.0 protective)
3. How strong is the statistical association? (size of OR/RR)
4. How sure are we that the association is truly causal? (Have we ruled out alternative explanations – confounding and chance?)

Randomized vs Non-randomized Exposure Groups

How is exposure determined?	Major class of design	Main types
Randomization (true computer-generated sequence, not synonymous with 'haphazard')	Experimental Randomized Clinical Trial (RCT)	Parallel group Crossover Factorial
Non-randomized • Self-selected (subject's choice) • Other-selected (parent, provider, insurance company) • 'Natural' (e.g. tornado)	Observational	Case-control Cohort Cross-sectional Ecologic

Observational

- Subject to more biases than RCTs
- Control of confounding variables can only be done for known and measured variables
- Can be more generalizable

Experimental

- Most rigorous:

causal evidence when available

- Randomization assigns known *and unknown* factors on the basis of chance
- Better on average:

Not infallible – individual RCTs can be ‘wrong’

Observational

- Can be used to directly study disease etiology
- May be only evidence for Treatment where randomization was not done first, or after widespread adoption
- May detect rare adverse events better

Experimental

- Cannot be used to study etiology (unethical to deliberately cause disease in humans)
- Only exception is “risk factor modification” studies (e.g. MRFIT, Gordis p. 155)

Main Observational Designs

Case-control

- Unmatched
- Group-matched
- Individual-matched
- Nested

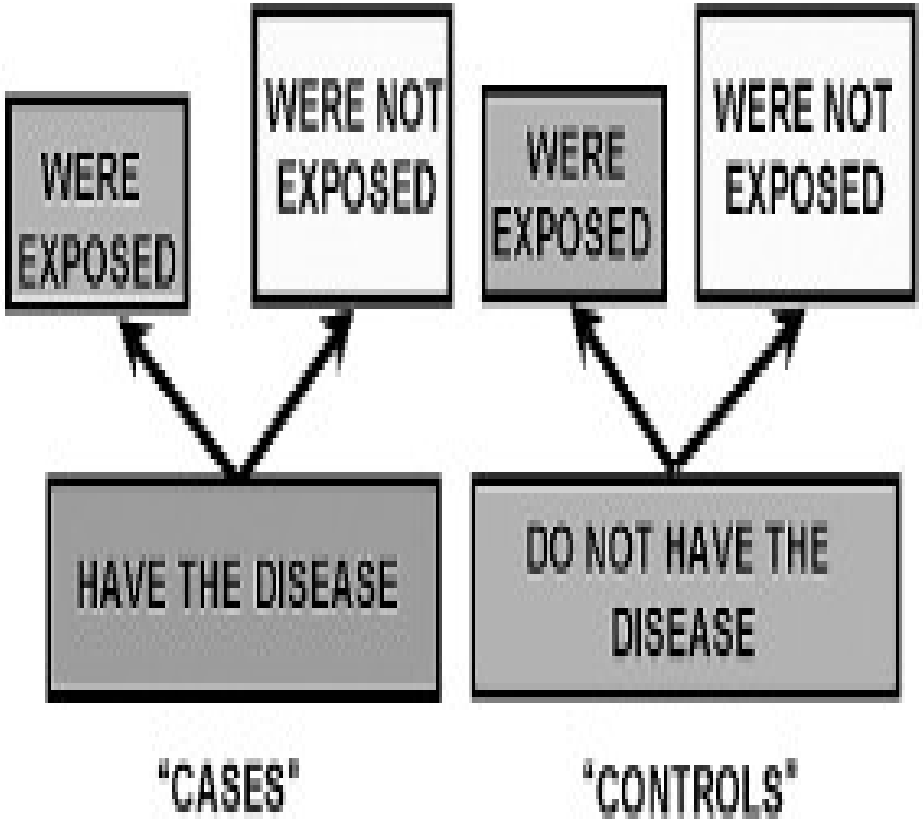
Cohort

- Retrospective
- Prospective

Plus hybrid designs for both …

Case-control Schematic

Step 1: Sample by Outcome



Step 2: Measure Exposure(s)

Case-control Studies

1. Select 2 separate samples on disease/outcome status:
   - CASES = have disease/outcome of interest
   - CONTROLS = no evidence of disease/outcome of interest

2. Measure exposure status in each group

3. Exposures and potential confounding variables are measured among cases and controls using same methods

4. Results expressed in 2 x 2 (or 2 x n) tables

Case-control Studies

- Advantages:
  - Very efficient
  - Can quickly identify strong effects
  - Study rare diseases and outcomes
  - Study multiple exposures

- NOTE: Since groups are defined on outcome, can only study a single outcome

- If cases and controls are representative of (population) cases and controls (see previous Slide):
  - higher odds of exposure among cases may indicate a causal risk factor (after controlling for major confounders)
  - lower odds of exposure among cases may indicate a causal protective factor (after controlling for major confounders)

Sources of Cases

- Diagnosed patients from hospital records, physician charts, tumour registries

- Single sources (e.g. one hospital) *may* cause bias through referral patterns (e.g. disease severity or type; physician speciality)

- Incident or prevalent cases used

NOTE: Prevalent cases may reflect factors affecting *survival*, not initial disease occurrence (remember the bathtub…)

Sources of Controls

Non-hospitalized persons:

- General population (door to door or telephone enumeration, voter lists, etc.)
- Neighbours, Friends, Siblings, Spouses

Hospitalized persons:

- Patient in next bed
- Patients from other wards or services
  - specified conditions
  - general patient sample

A case-control study of ginseng intake and cancer

Int J Epidemiol 1990;19:871-876

- Cases n = 905
  - 'Newly diagnosed' pathologist confirmed cancers between Feb 1/87-Jan 31/88
  - Aged 20+
  - Excluded ICU and otolaryrngology admissions for interview reasons
  - Single major cancer hospital in Seoul (14% of all cancer diagnoses)
- Controls n = 905
  - Non-cancers from same hospital
  - Matched by gender; age at Diagnosis (+/- 2 yr); date of admission (+/- 3 mo)
  - Excluded diseases "associated with smoking or alcohol consumption" but included cardiovascular, COPD, peptic ulcer, and liver cirrhosis

Measurement Information from Study

- Two trained interviewers
- Same case-control pair per interviewer
- Demographics collected
  - age, marital status, education, occupation, income
- Lifestyle information collected
  - tobacco and alcohol use
- Ginseng consumption using dietary recall to examine “lifetime” use
  - Ginseng classified as fresh, white and red, and form (sliced, extract, powder, etc.)
  - Types/products, frequency (daily, monthly) for each decade of life

Selected Results

Comparability of cases and controls reported in Table 3:

	Case%	Control%
– Married	88.1	88.8
– Professional	3.2	4.6
– Agricultural	24.0	19.0
– Low income	27.1	30.6
– No education	17.2	17.6
– College	5.8	7.3

Ideal in all studies is for balance on all factors not of primary interest

Results (Observed)

Ginseng	Cases		Controls	
Any	562	a	674	b
None	343	c	231	d

Odds ratio = (a x d) / (b x c) = (562 x 231) / (674 x 343) = **0.56**
95% Confidence interval (0.46 – 0.69)

Ginseng was associated with a (100% - 56%) = 44% reduction in cancer
(in line with the RR of 0.49 from the prospective study, Class 2)

Two Major Sources of Bias

I. Sampling

- Unrepresentative cases and/or controls
- selection bias causing selection of cases and/or controls to be more or less exposed than in the population of interest

ii. Measurement

- Accuracy of exposure measurement differs between cases and controls
- Accuracy of outcome measurement differs between exposed and unexposed

Sampling Bias

How does it work?

"Wanted: Women with silicone breast implants for a study of unexplained health problems associated with breast implants."

- Assume 5% prevalence of breast implants
- Assume 5% prevalence of unexplained health problems
- Assume H0 is true

Which groups might be more or less likely to volunteer for a study like this?

Hypothetical Results in 10,000

Assume true population distribution

Exp	Case	Control
+	25	475
-	475	9025

True OR = 1.0

Assume biased sample distribution

Exp	Case	Control
+	23 (92%)	238 (50%)
-	238 (50%)	4061 (45%)

Biased OR = 1.6 (1.03 – 2.5, p = 0.037)

Percentages are participation rates of people in each cell
e.g. Cell a, 23/25 = 92%

Measurement Bias

- Recall Bias:
  - "Systematic error due to differences in accuracy or completeness of recall to memory of past events or experiences. For example, a mother whose child has died of leukemia is more likely than the mother of a healthy living child to remember details of such past experiences as the use of x-ray services when the child was in utero."

Porta, 2008:208

Measurement Bias
(specific example: recall bias)

How does it work?

Always assume H0 is true (OR = 1.0) and think of the observed 2 x 2 table:

- If cases *more likely* to report exposure (and/or controls are *less likely*), a and/or d will be larger than truth and OR will be biased > 1.0
- If cases *less likely* to report exposure and/or controls *more likely*, a and/or d will be smaller than truth and OR will be biased < 1.0

A case-control study of pesticides and fetal death due to congenital anomalies

Epidemiology 2001; 12:148-156 (coursepack)

- Cases: n[1] = 73, from birth, fetal death and death certificates from 10 rural counties in California

- Controls: n[1] = 611 live normal births, randomly chosen from same 10 counties, matched by county of mother's residence and mother's age (+/- 5 yrs)

[1]Note: case n does NOT have to equal control n

Potential Confounders in Study

- Race/ethnicity
- Gender of fetus
- Trimester prenatal care began
- Season of conception
- Prior fetal loss

No major imbalances between cases and controls were found in the study

Exposure Assessment in Study

- California Pesticide Use Report:
  - Specific chemicals, amount, date, location
- Maternal address and last menstrual period (LMP) from fetal death, death and birth certificates
- Maternal address linked to pesticide use within 1 square mile (narrow definition) or within adjacent 8 square miles (broad definition)
- Permitted detailed exposure assessment for 327 chemicals for each day of each pregnancy (estimated days since conception = LMP + 14 days)

Study Results

- Largest risks seen for exposures 3rd – 8th week of pregnancy

- Pesticide Class ORs:

Phosphates, carbamates, endocrine disrupters	1.4 (0.8,2.4)
Halogenated hydrocarbons	2.2 (1.3,3.9)

Results consistent under different methods of exposure assessment (same square mile, adjacent 8 square miles)

Strengths and Limitations of Study

- No potential for recall bias
- Vital statistics reasonably good, no reason to expect bias by pesticide use
- Consistent pattern of increased OR under alternative analyses
- Biological basis for hypotheses (period of organogenesis)
- Mothers may not have lived at their recorded address during pregnancy
- No biological samples
- Finest detail is up to 1 mile
- No meteorological data to link airborne spraying to wind direction

Recall Bias: Major potential bias in Case-control Studies

- Recall bias:
  - Accuracy of exposure measurement differs by case-control status due to biased recall
- Solutions:
  - Use other measures than self-report (proxies, official records)
  - Blind participants to specific hypotheses (e.g. ‘lifestyle factors’ rather than ‘smoking’)
  - Blind interviewers to case-control status
  - Standardized interview

MacMahon, 1981 (Gordis p. 183)

- Cases had pancreatic cancer
- Controls were selected by cases' physicians
- Many patients with pancreatic cancer are admitted by gastroenterologists, who also treat stomach disorders
- An exposure of interest was coffee drinking
- How may this have biased the study?
- How could you test this with study data?

Hospital vs *Population-Based* Control Subjects in Case-control Studies

<u>Hospital</u>

- Convenient
- Usually higher response rate
- Should include various diagnosis groups so no particular risk factors are overrepresented

<u>Population</u>

- Increasingly common due to administrative data on entire populations
- May be more representative of exposure in the population

Case-Control Matching

- Why?
  - To increase comparability between groups on important factors other than primary exposure of interest
- How?

1. ***Group*** or ***Frequency*** Matching: Match two groups on key characteristics (e.g. cases and controls both have 53% female). Requires selection and descriptive analysis of cases before controls selected.

2. ***Individual (matched pairs)*** Matching: Match each individual (e.g. each 25 year old female case matched with the next eligible 25 year old female control). Often used with hospital controls.

Limitations of Matching

- Practicality:
  - Severe limitation as number of matching variables increases
- Conceptual:
  - Cannot analyze effect of matching variables. Why?
- Overmatching*:*
  - Needlessly controlling for extraneous factors
- Best case-control studies only match on strong (already established) risk factors that are not of interest as exposures in that particular study

Varying the Ratio of Cases to Controls

- Simplest case-contrl study has 1:1 (ginseng example)
- With rare diseases/outcomes the number of cases is limited (fetal anomalies example)
- Study power (probability of finding a causal relation if one exists) increases with more controls (1:2, 1:3, etc.)
- The increase in power diminishes after 1:4
- With any ratio ≥1:2, opportunity to use different sources of controls:
  - Hospital (Control Group1) and Community (Control Group 2)
  - Disease (or ward) A and Disease (or ward) B

→ If Odds Ratios are similar when calculated separately for the 2 sources of controls, have evidence that choice of control group is not introducing a bias.

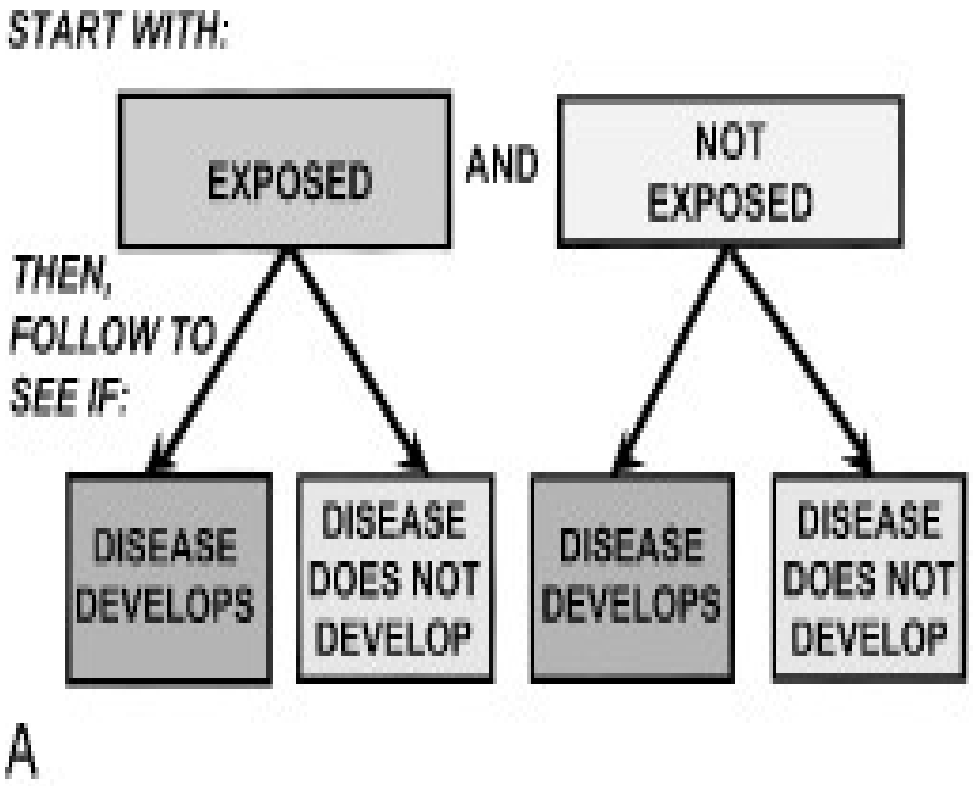
→ The 2 control groups can be combined to increase statistical power and precision.

How Efficient is the Case-control Design?

True OR in Population	# Case-control Pairs	
10	22	Extremely efficient for large effects
8	27	
4	67	
2	307	Not bad even for small effects
1.5	957	

To have an 80% chance of detecting an OR this large or larger, at the 95% confidence level, with 1:1 case-control ratio, assuming the frequency of exposure in the controls is 10% (can be done in OpenEpi)

Cohort study schematic



Cohort studies

Initial focus on *exposure*, not *outcome* (disease, injury, etc.)

Example 1: ***Classical*** method heightens *exposure contrast*; e.g. many studies of occupational exposures:

Exposed: asbestos miners

Unexposed: other miners

Example 2: ***Community*** method heightens exposure representativeness; e.g. Framingham:

Select people from a general population who will naturally vary on several exposures e.g. smoking, alcohol, diet, exercise, etc.

Classical

Exposure examples:

- Occupational (chemicals, stress)
- Residential (toxic dump sites)
- Procedures, programs and treatments (abortion, day care, thalidomide, DES *in utero)*

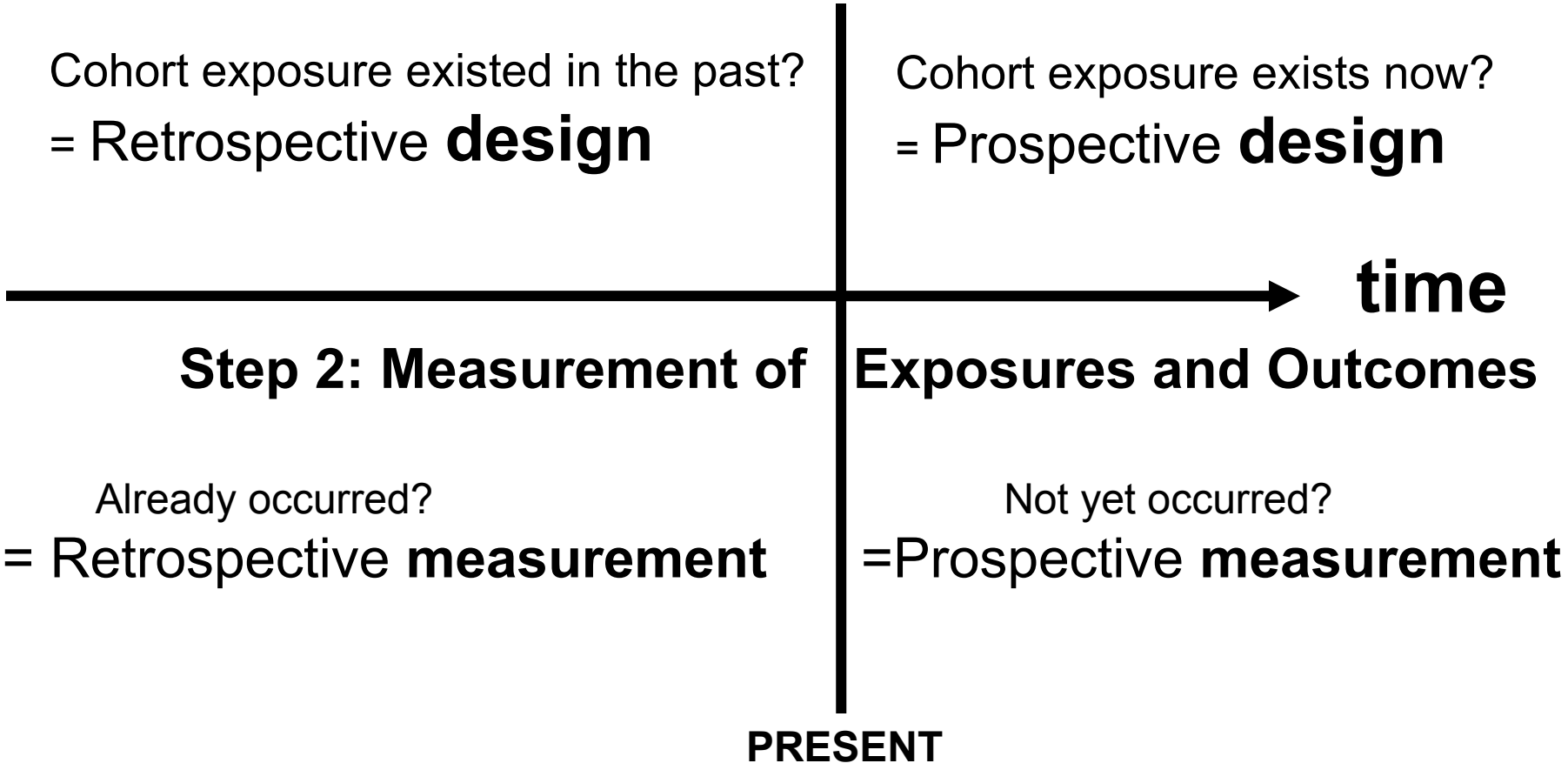
Community

Exposure examples:

- anything measured at baseline can then be considered for analysis (past occupations, diet, tobacco/alcohol/other drugs)

Retrospective – Prospective
Design vs. Measurement

Step 1: Study design – define exposure cohorts of interest



Two cohort studies of the effect of being exposed to E2200b on risk of becoming an epidemiologist, started today:

- Exposure definition: Western students (current 2200b vs. non-2200b)
  - Design: Prospective Cohort Study
  - Measures:
    - Retro: Entering GPA, Average UWO grades, previous research experience, other methods courses, etc.
    - Pro: MSc Epi program after graduation

- Exposure definition: (E3330b students/non-3330b)
  - Design: Retrospective Cohort Study
  - Measures:
    - Retro: Entering GPA, UWO grades, previous research experience, MSc in Epi (after graduation, employment as 'Epidemiologist'
    - Pro: PhD Epi program, employment as 'Epidemiologist'

Non-organ specific cancer prevention of ginseng: a prospective study in Korea

Int J Epidemiol 1998;27:359-364 (coursepack)

- Sampling frame: provincial residents' list 1987
- Born < 1947 (aged 40+)
- Standard questionnaire w/ trained interviewers, total n = 4,634:
  - lifetime occupations, smoking and alcohol, past diseases
  - Ginseng: ever consumed, age 1st consumption, frequency, duration, type.

100% follow-up for cancer and non-cancer deaths though Dec. 1992

Risk Ratios for ‘dose response’

(one criterion used to assess causation)

Ginseng	# Cancer	# At risk	5 yr Risk	Risk ratio*	(95%CI)
None	62	1345	0.046	1.0	--
1-3 / yr	39	1478	0.026	0.56	(.37 - .86)
4-11/ yr	21	945	0.022	0.47	(.30 - .79)
Monthly +	15	818	0.018	0.40	(.23 - .70)

*Each risk ratio calculated with risk for ‘none’ (“Reference Category”) in the denominator:

1-3 / yr = 0.026 / **0.046** = 0.56

4-11 / yr = 0.022 / **0.046** = 0.47

There is a successive gradient in reduced cancer risk with increasing frequency of ginseng consumption; each stratum is stat. significant

Strengths and Limitations

- Logistic regression models adjusted RR for age, gender, education, smoking and alcohol
- 100% follow-up
- Consistent effect for lung and stomach cancer
- Replicates earlier case-control study
- Consistent with animal models
- Generalizability: primarily agricultural population from ginseng farming area
- 5 year follow-up too short to have enough cases of rarer cancers for analysis
- No control for known or suspected dietary and viral confounders

Poundstone, AIDS 2001: Retrospective cohort design

- Cohorts were HIV-infected patients defined on the basis of when they were treated:
- Era 1, 1990-1995 (Pre-HAART)
- Era 2, 1996-1999 (Post-HAART)

While this is a study of treatment effectiveness rather than risk factors, it uses retrospective cohort methods.

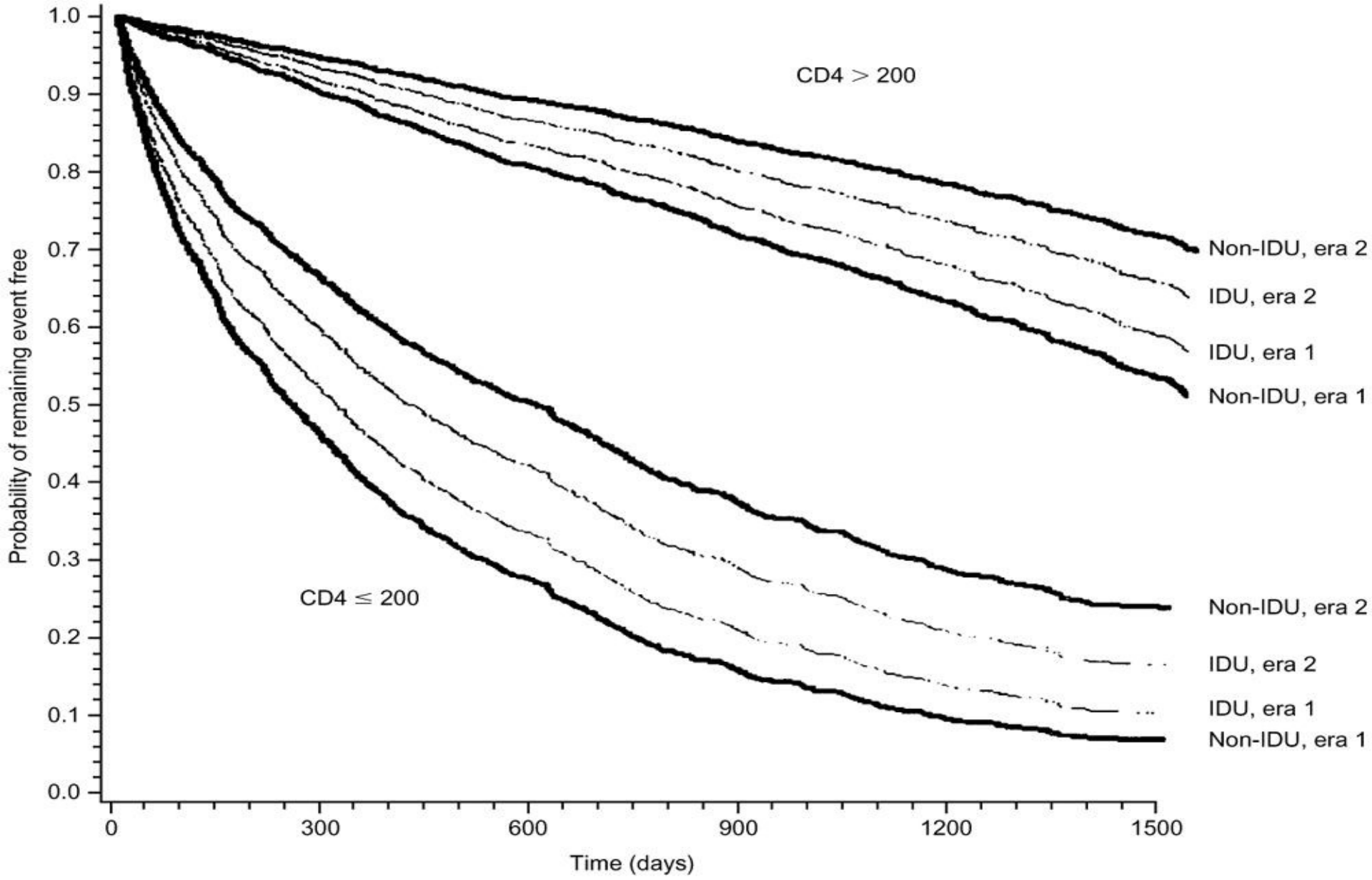


Fig. 2. Probability of remaining event (new AIDS-defining illness or death) free over time stratified by CD4 cell count (<= 200 versus > 200 cells/mm3), and by injecting drug use (IDU) versus non-IDU HIV transmission risk, and era 1 versus era 2. *From:* Poundstone: AIDS, Volume 15(9).June 15, 2001.1115-1123

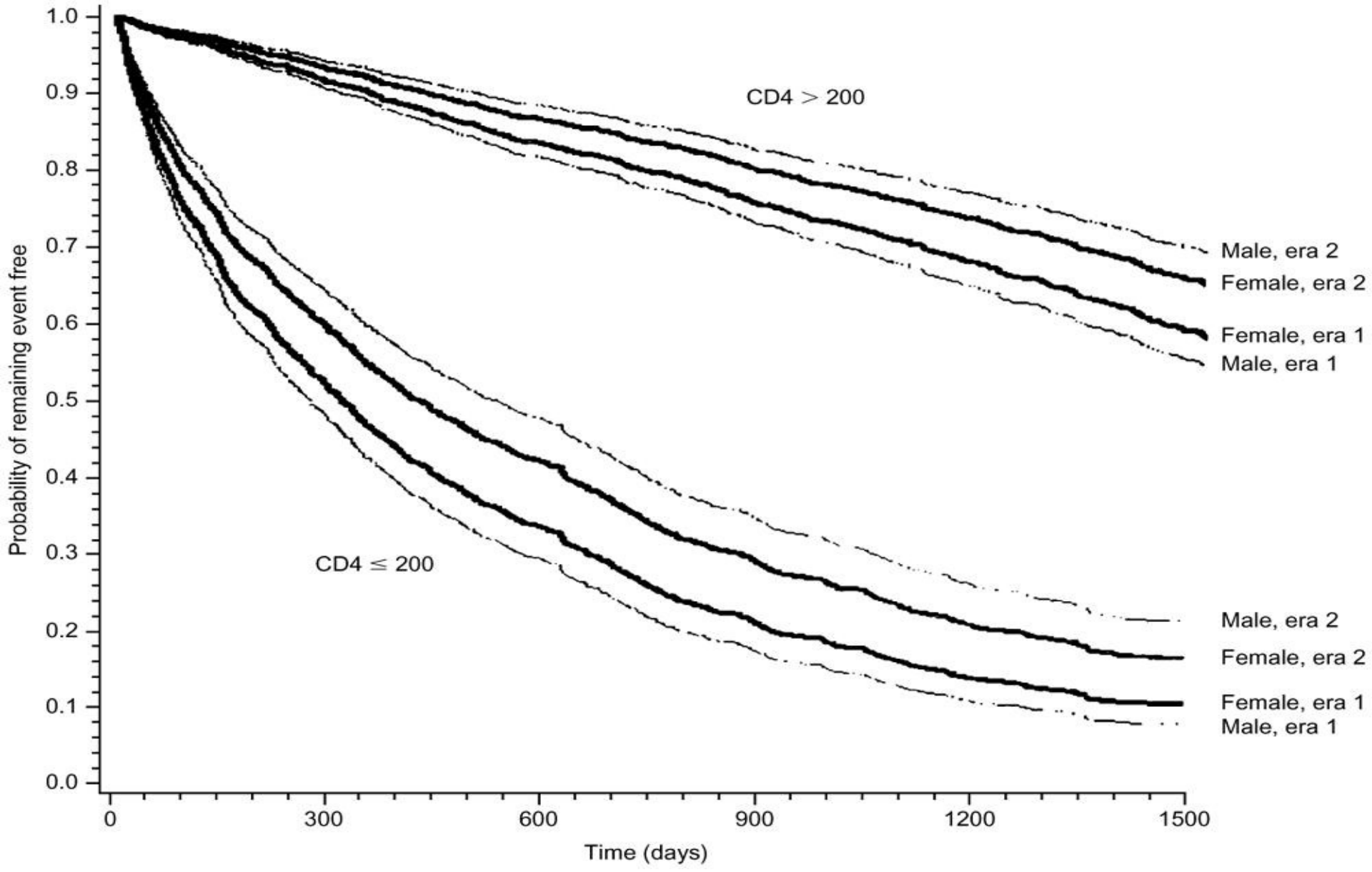


Fig. 3. Probability of remaining event (new AIDS-defining illness or death) free over time stratified by CD4 cell count (<= 200 versus > 200 cells/mm3), and by sex (male versus female), and era 1 versus era 2. *From:* Poundstone: AIDS, Volume 15(9).June 15, 2001.1115-1123

Observational studies of treatments

- Results show clear differences in AIDS-free survival time in Post-HAART era
- Why did she stratify by CD4 count, IV drug use, and gender?
- Why did she not use an RCT?
- Would you approve an RCT today?
- Would you choose HAART based on these non-randomized findings?
- Would you participate in an RCT if you were HIV+, based on Poundstone's result?

Remembering Cohort Designs

- Cohort designs always have study groups defined on the basis of **exposure**, either in a high/low exposure contrast study (e.g. different occupations), or (hopefully) representative of exposures of primary interest in the population to whom you wish to generalize (e.g. Framingham, Gordis p. 171)
- Can study multiple outcomes for given exposure

Major potential biases in cohort studies

Gordis p. 174

Bias Name	Mechanism	Solutions
Assessment of Outcome	Assessors know exposure group and study hypothesis	Blind assessors Test for awareness
Information	Quality and/or extent of information differs by exposure group	Prospective: identical data protocol Retro: test hypotheses using equivalent data
Non-response and loss-to-follow-up	Participants/non participants have different characteristics Those lost have different outcomes than those tracked	Maximize response and follow-up rates Compare groups using known information
Analytic	Strong preconceptions among analysts	A priori analysis plan re: primary outcomes

Cohort study sample sizes

True Risk Ratio	Disease incidence in unexposed (%)	Required sample sizes Exposed	Unexposed
10	1	83	333
	5	12	47
8	1	113	452
	5	17	69
4	1	335	1,339
	5	59	236
2	1	1,673	6,691
	5	313	1,253
1.5	1	5,305	21,219
	5	1,006	4,026

80% chance of finding RR this large or larger, with 95% confidence, assuming a 25% risk factor prevalence

Choice of Design: Optimal conditions

Case-control

- Outcome (disease) is “rare” (but measurable)
- Relatively little is known about risk factors
- Promising hypotheses of specific risk factors (may be from cross-sectional or case series observations)
- Answer needed quickly (emergency)

Cohort

- Exposure is common and well defined
- Sufficient number of outcomes will occur during study period
- Some evidence exists from case-control studies
- Reliable records exist (retrospective cohort)

Class Discussion: Design Choices

How would you study the following questions:

1. Is fluorescent lighting a risk factor for migraine headache?
2. Is ‘examination stress’ a risk factor for colds and flu among students?
3. Are cosmetic pesticides a risk factor for cancer in children?

Design choices (continued)

4. Do modern laboratory fume hoods reduce the risk in chemistry majors of disorders caused by common laboratory chemicals?
5. Is recreational hockey a risk factor for myocardial infarction in middle-aged men?
6. In middle-aged women?

(Research questions contributed by the class)

Summary of Strengths
(see p. 225 of Gordis)

Case-control

- useful for rare diseases
- useful for long latency or induction periods
- relatively inexpensive
- relatively quick
- multiple exposures in one study

Prospective Cohort

- risk factors measured before outcome known, X – Y temporality known with certainty
- multiple outcomes in one study
- yields incidence estimates

Summary of Limitations

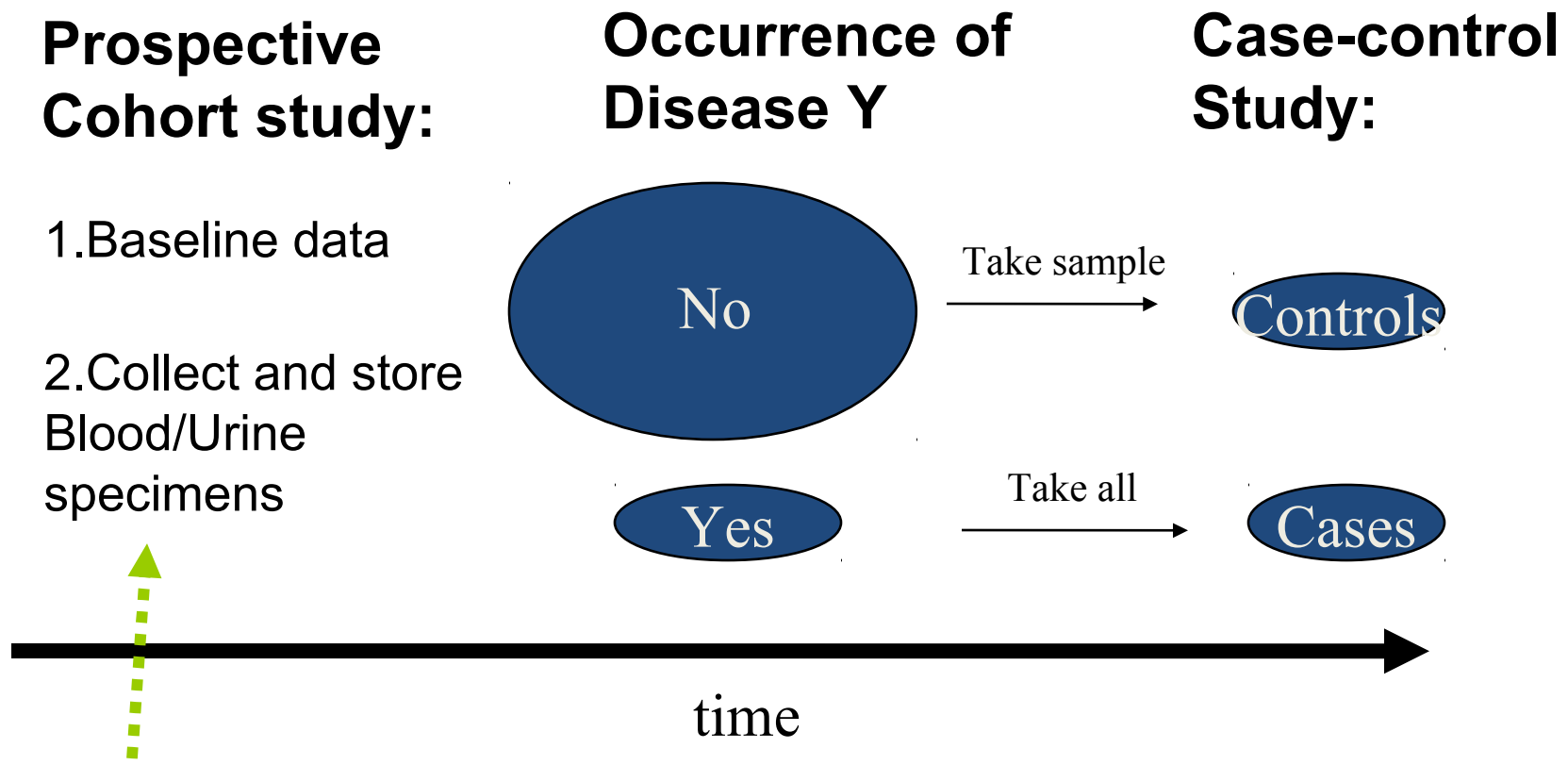
Case Control

- potential for recall bias in measuring risk factors when disease known
- cases may not represent population with disease
- controls may not represent population without disease

Prospective Cohort

- expensive
- can require large sample
- can require long follow-up
- Biases: (Gordis, p. 156)
  - Assessment of outcome
  - Information bias
  - Non-response and loss to follow-up
  - Analytic bias

A hybrid design:
Nested Case-Control study



Intensive analyses of fewer specimens saves cost (increases efficiency)
Temporality assured; baseline data free of recall bias.

Nested Case-Control study

Meat intake and risk of stomach and esophageal adenocarcinoma within the European Prospective Investigation Into Cancer and Nutrition (EPIC).

J Natl Cancer Inst. 2006 Mar 1;98(5):345-54.

Begins with cohort: n=521,457 aged 35-70

Baseline dietary and lifestyle questionnaire

Mean Follow-up 6.5 yr

Adenocarcinoma Cases: 330 gastric, 65 esophageal

Controls: No evidence of above

Results:Gastric cancer risk was statistically significantly associated with intakes of total meat (calibrated HR per 100-g/day increase = 3.52; 95% CI = 1.96 to 6.34), red meat (calibrated HR per 50-g/day increase = 1.73; 95% CI = 1.03 to 2.88), and processed meat (calibrated HR per 50-g/day increase = 2.45; 95% CI = 1.43 to 4.21).

Wanna grab some
4-sausage pizza?

- Nested Case-Control assess exposures before outcome; no recall bias (as with prospective cohort studies) combined with efficiency of Case=Control studies
- Many potential confounders were controlled :

Cross-sectional design

- Current existence of an outcome is correlated with current existence of an exposure in 1 sample

(Note the difference with case-control studies which measure *past* exposures in 2 samples)

- Problems as analytic design:
  - *temporality*: is X a result of Y, or the other way around, or are both affected by another factor?
  - Ignores natural history (incident vs. prevalent cases)

Ecologic Studies

Exposures and outcomes measured in **populations** not individuals (e.g. a positive correlation has been noted between per capita fat consumption and breast cancer rates across several different countries)

- Easy to do but suffer from several potential biases (e.g. *cross-level bias* would occur if individual women with breast cancer actually had low fat diets, and vice versa.
- Because there is no way to control this bias, ecologic studies are best for hypothesis-generation.

Cross- sectional, Case-control, or ecologic? Identify the bias(es)

1. Mothers of premature infants report more stressful events during pregnancy than mothers of full-term infants.
2. The prevalence of obesity and of depression are positively correlated among 500 high schools.
3. Obesity and depression are positively correlated in a sample of 500 high school students.

- Design a better study for each causal question

Sample exam question #1

To assess the association between mercury dental fillings ("amalgams") and Multiple Sclerosis (MS), investigators select 200 subjects with MS and 200 subjects without MS, and interview each about amalgams received since childhood. This most closely resembles:

A) retrospective cohort

B) prospective cohort

C) case-control

D) cross-sectional

E) ecologic

Hint: Think about how the groups were assembled – disease status or exposure status?

Sample exam question #2

To assess the association between alcohol consumption and heart disease, investigators select 1000 people, 700 who drink alcohol regularly and 300 who drink infrequently or not at all. None have heart disease at baseline. Subjects are followed for 10 years and new cases of heart disease are recorded in each group. This most closely resembles:

A) retrospective cohort

B) prospective cohort

C) case-control

D) cross-sectional

E) ecologic

Sample exam question #3

To assess the association between exposure to paint fumes and brain cancer, employer records from 1960 – 1980 are used to identify people who worked as painters, and a second group who did not work as painters. Insurance claims from 1960 to the present are searched for cases of brain cancer in the two groups. This most closely resembles:

A) retrospective cohort

B) prospective cohort

C) case-control

D) cross-sectional

E) ecologic

Sample exam question #4

- In a case-control study of therapeutic abortion and breast cancer, self-reports of abortion histories are compared to medical records in both cases and controls. The cases are found to over-report, and the controls to under-report, their exposure to abortion. If the self-reported abortion history were used in the analysis, it would result in:

A) information bias

B) non-response bias

C) referral bias

D) recall bias

E) cross-level bias

Sample Exam Question # 5

- To study the relationship between exposure to glass dust and respiratory disease, employee and hospital records from Empire Glass and Consolidated Wood Products workers are compared. Empire Glass has had regular respiratory health assessments for the whole follow-up period; CWP has had none. This could cause:

A) recall bias

B) analytic bias

C) non-response bias

D) Information bias

E) cross-level bias

References and Further Reading

- Rothman KJ. Epidemiology: An Introduction. Oxford Univ. Press, 2002.
- Kelsey JL, Whittemore AS, Evans AS, Thompson WD. Methods in Observational Epidemiology, 2nd ed., New York: Oxford Univ. Press, 1996.
- All sample sizes, OR/RR and 95% CI performed using www.OpenEpi.com