Case-Control Studies

- Compare Diseased with Not Diseased on Previous Exposures
- “aims to establish the relationship of cases to antecedent factors in a retrospective manner”
- Instead of looking at the probability of disease given exposure, look at the probability of exposure given disease
- Hill and Doll studies of lung cancer and smoking

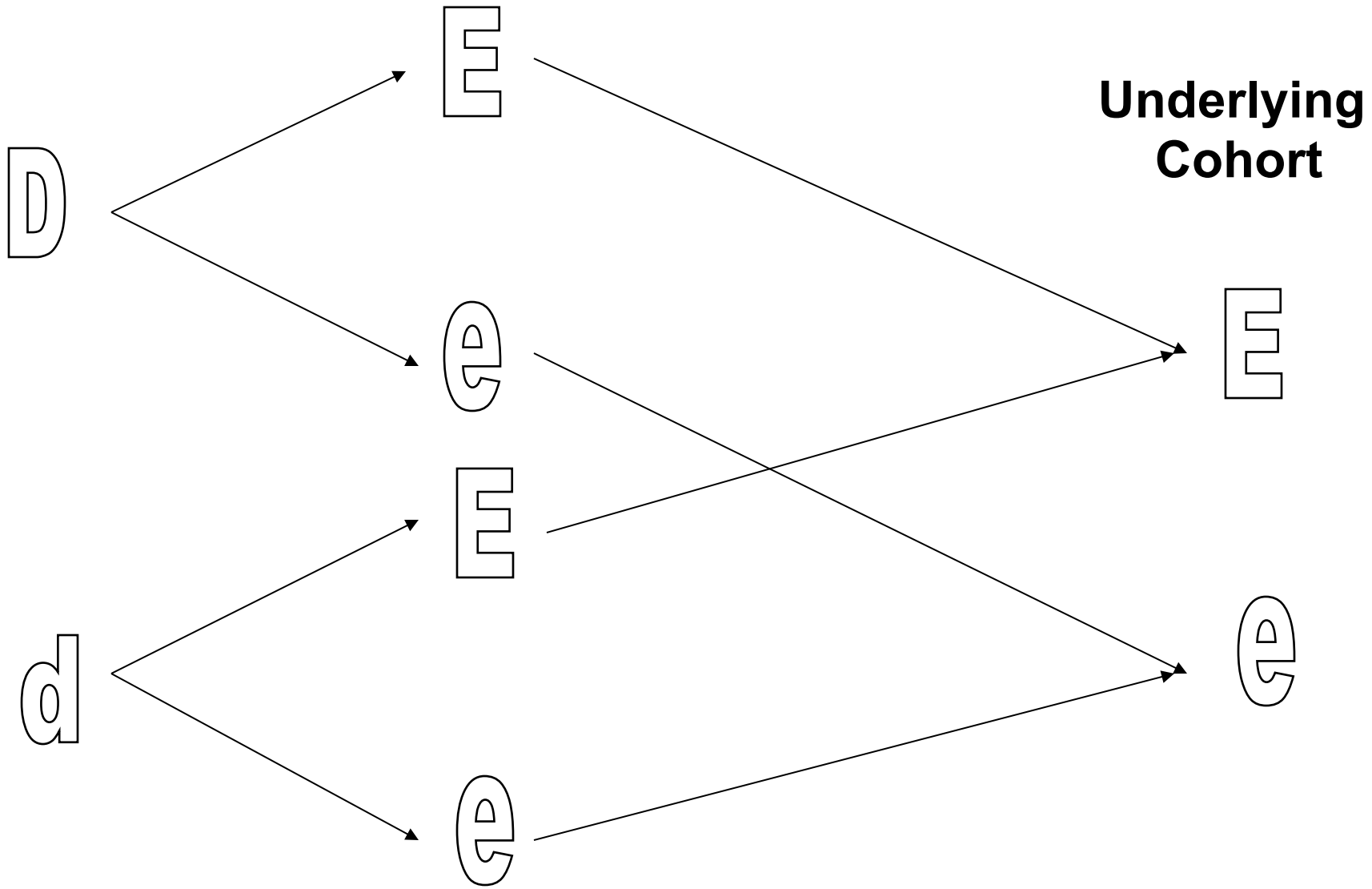
Advantages

- Cost
- Time
- Rare Diseases
- Diseases with long latency periods
- IDs (CDC)

Disadvantages

- Temporality
  - Did exposure actually precede disease?
  - Difficult to quantify level of exposure
  - Better if rapid onset disease
- Control Group – crux of the problem
  - “the control series is intended to provide an estimate of the exposure rate that would be expected to occur in the cases if there was no association”
  - ***study base*** “the most frequently used source of controls is people seeking care at the same (hospital) for other diseases”
- Recall Bias

Anatomy of a Case-Control Study



Analysis of Case Control Studies: The Odds Ratio

- Prospective vs. Retrospective Approach
  - Cohort studies: Pr[D|E] e.g. Pr[CA|Smoking]
  - Case-control: Pr[E|D] e.g. Pr[Smoking|CA]

Are they measuring the same thing?

Smoking and Lung Cancer

LUNG CANCER				
SMOKING		**Yes**	**No**	
	Yes	**100**	**900**	**1000**
	No	**50**	**1950**	**2000**
		150	**2850**	**3000**

$$\mathbf{Pr[D|E] = 100 / 1000 = 0.10}$$

$$\mathbf{Pr[E|D] = 100 / 150 = 0.66}$$

Need for a New Measure of Effect

- Recall: Odds related to Probability (Risk)
  - Odds = Probability/1 – Probability (And Probability = Odds / 1+ Odds)
    - 1:1 transformation; W = odds of A occurring, then p= P[A] = W / W +1, e.g. if odds = 2:1, probability = 2/3; if the probability = 0.75 (3/4) then the odds = (3/4) / (1/4) = 3:1

- ODDS = Pr[D] / Pr[d] = Pr[D] / 1 – Pr[D]

- ODDS RATIO $= \dfrac{\text{Odds in Exposed}}{\text{Odds in Unexposed}}$

A way for us to get at risk retrospectively…

Calculating The Odds Ratio

- OR = ad/bc
- Lung CA example, OR = (100)(1950) / (900)(50) = 5.0

	D	d
E	100	900
e	50	1950

- RR= 100/1000 / 50/2000 = 4.0

Derivation and Invariability of the Odds Ratio

- Exposure Odds Ratio (Pr E|D / PrE|d)
  - P[E | D] / P[e | D] = P[E | D] / 1 - P[E | D] = (a/a+c) / (c/a+c)
  - P[E|e] = P[E | d] / P[e|d] = (b/b+d) / (d/d+c)
  - OR = [(a/a+c) / (c/a+c)] / [(b/b+d) / (d/d+c)] = (a/c) / (b/d) = **ad/bc**
- Disease Odds Ratio (Pr [D|E] / Pr[D/e])
  - P[E | D] / P[e | D] = P[E | D] / 1 - P[E | D] = (a/a+c) / (c/a+c)
  - P[E|e] = P[E | d] / P[e|d] = (b/b+d) / (d/d+c)
  - OR = [(a/a+c) / (c/a+c)] / [(b/b+d) / (d/d+c)] = (a/c) / (b/d) = **ad/bc**

Rare Disease Assumption

- *The OR will approximate the RR if the disease is "rare"*
- Few people die from D, don't contribute much P-Y to denominator
- 'a' cell small relative to 'b'; 'c' small relative to 'd'
- RR = (a/a+b) / (c/c+d) ~ (a/b) / (c/d) = ad/bc = OR

	D	d
E	A	B
e	C	D

Cross-Sectional Studies

- All there was at time of epidemiologic transition
- Exposure and disease ascertained simultaneously; individual level data
- Inexpensive and simple
- Problems and Biases
  - Directionality
  - Incidence – Prevalence Bias
    - E.g. mouthwash and oral CA
  - Recall Bias

Evans County, GA.

	CORNOARY ARTERY DISEASE	NO CORONARY ARTERY DISEASE	TOTAL
PHYSICALLY ACTIVE	14	75	89
NOT PHYSICALLY ACTIVE	3	87	90
TOTAL	17	162	179

Relative Risk = (14/89) / (3/90) = 4.7

Problems and Biases

- Directionality
  - Mouthwash and Oral CA
  - Hip Fx and Obesity
  - CAD and Activity
- Incidence – Prevalence Bias
  - More likely to pick up chronic cases
  - Evans County: CAD Prevalence higher in whites vs. blacks
- Recall Bias
  - Birth defect studies

Ecologic Data vs. Individual-Level Data

- A. Ecologic Studies (proportions, percentages)
  - Advantage – cheap, easy, fast, new hypotheses, to study group-level attributes
  - Problem – ecologic fallacy
- B. The Ecologic Fallacy
  - Aristotle's "fallacy of division
  - " the assumption that an association at one level of organization can be inferred from that at another"
  - "cross-level" analysis
  - E.g. Durkheim, Robinson, Lung Cancer and pollution

We don't know the cells, only the marginals:

	Disease	No Disease	Total
Exposed	?	?	A+B
Not Exposed	?	?	C+D
Total	A+C	B+D	A+B+C+D = total

Ecologic Fallacy

- Durkheim
  - Suicide rates in Prussian provinces strongly correlated to proportion of Protestants (8X ↑)
  - Individual data➔ risk ↓ to 2X
- Robinson
  - Literacy
  - r=0.62 areas with many recent immigrants

Design Features of Ecologic Studies

- Unit of Analysis the group (often defined geographically)
- Data more readily available
- Inexpensive, quick, can generate useful hypotheses
- Often only way to study group-level variables
- Correlations often much higher than those seen in individual-level studies
- Does disease occur in exposed? (fallacy)