

The Chamber of Epi “Secrets”: Epidemiologic terms and nomenclature

PulseNet Update Meeting

April 28, 2004

Michael Lynch, MD, MPH

**Foodborne and Diarrheal Diseases Branch
Centers for Disease Control and Prevention**



Objectives

- **Review epidemiologic terms commonly used by epidemiologists in outbreak investigations**
 - **To get connected we need to speak the same language**

- **Review Harry Potter’s investigation of an outbreak of “petrification” at Hogwart’s School**
 - **Several students turned to stone (“petrification”)**
 - **Foul play suspected**
 - **Food exposures overlooked**

A case is a case

- A case is a particular instance of a disease, health disorder, or condition under investigation
 - diarrhea
 - salmonellosis
 - “petrification”
- A case is not the person with the disease (though we often speak of them like they are people)
 - patient or case-patient refers to the person

The case definition is what we say it is

- The characteristics (time, place, person and clinical features or symptoms) of the case being studied

Petrification: “an instance of complete, tonic, immobilization with onset between April 27 and 28 in a person residing at Hogwarts’s School of Wizardry”

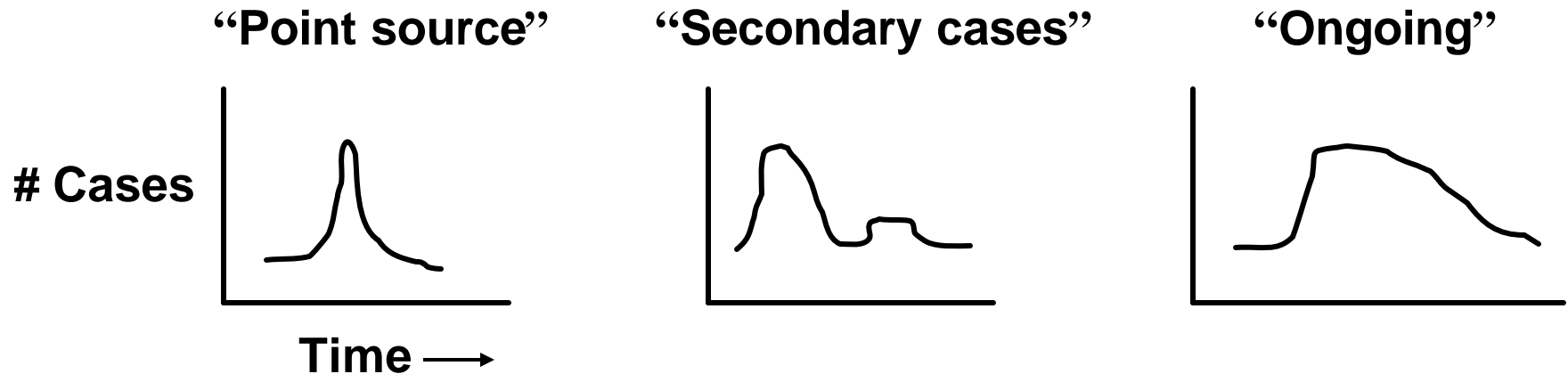
- Can vary from broad (non specific) to narrow (highly specific)
- Suspect, probable, confirmed case

Is it a cluster, an outbreak, or an epidemic?

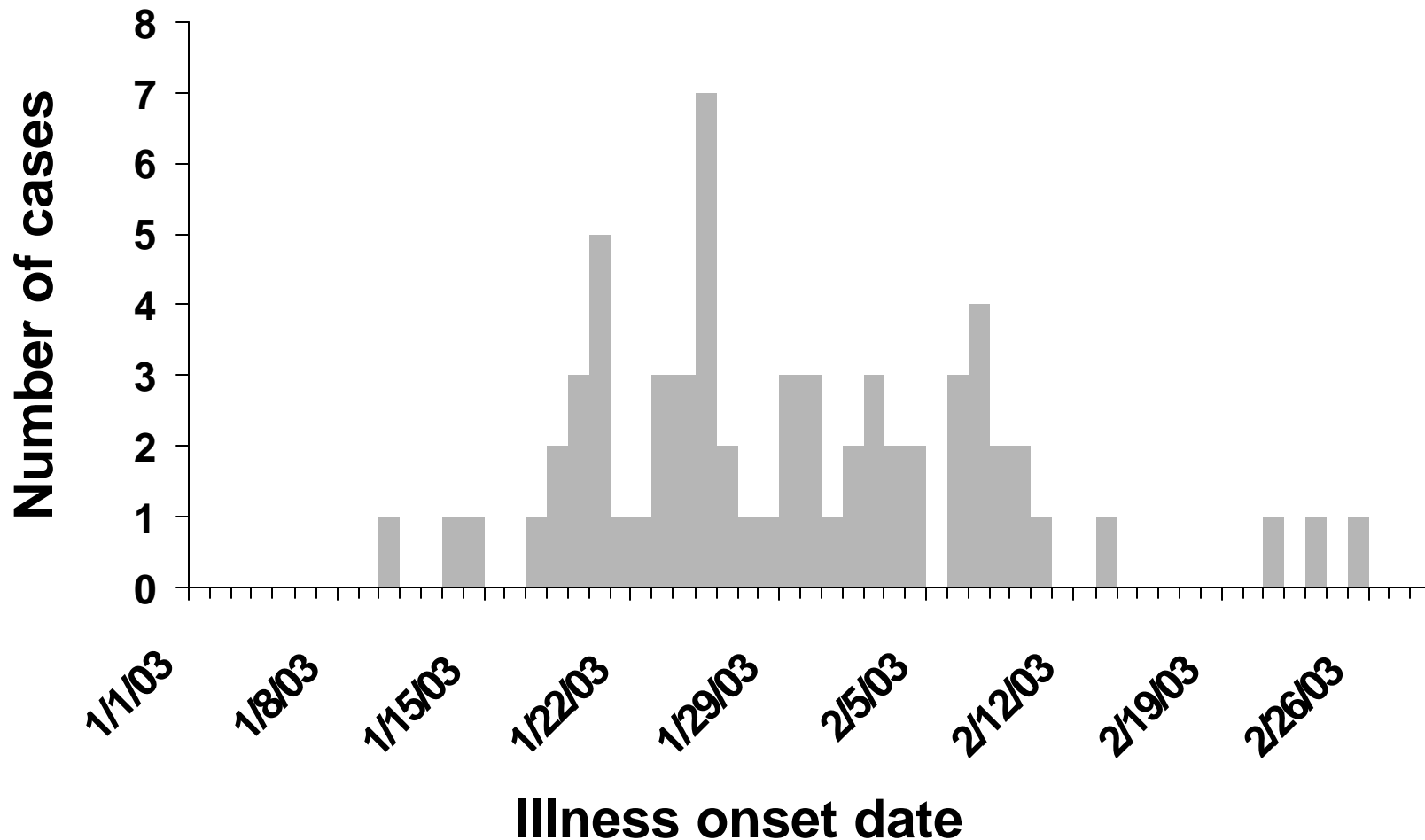
- **Cluster** is an aggregation of relatively uncommon events or diseases in time and space in amounts that are believed or perceived to be greater than expected by chance
- **Outbreak** is an increase in the expected incidence of disease, limited in time and/or area (with a presumed or proven common source)
- **Epidemic** is an increase in the expected incidence of disease, over prolonged time and/or large area

We use descriptive epidemiology to describe the outbreak

- The who, what, when and where of the outbreak
- Age, gender, illness onset date, location
- Epicurve plots cases in time



***Salmonella* Newport* multi-state outbreak cases by illness onset date, United States, January–February, 2003 (n=68)**



* PulseNet PFGE pattern JJPX01.0245

Coming up with ideas about what happened: Hypothesis generation

- Hypothesis: A supposition arrived at from observation or reflection that leads to refutable predictions
- We interview ill persons to generate hypotheses about cause of outbreak
 - Open-ended and closed questions
 - Demographics, food/beverages, travel, activities

Exposures, vehicles and reservoirs

- **Exposure**: Proximity and/or contact with a source of a disease agent
- **Vehicle of transmission**: the mode of transmission of an infectious agent from its reservoir to a susceptible host
- **Reservoir of infection**: Any person, animal, plant, soil, or substance, or a combination of these, in which an infectious agent normally lives and multiplies

Analytical studies test hypotheses

- **Evaluate the association between exposure (E) and disease (D)**
- **Two used in foodborne outbreak investigations**
 - **Cohort study, subjects identified based on exposure status**
 - **Case-control study, subjects identified based on case status**

Cohort study

- **Identifiable & manageable setting**
 - Single event (e.g. church supper)
 - Single location (e.g. cruise ship)
- **Assess risk of illness if exposed**
- **Identify which exposures are associated with disease**

Case-control study

- **Unclear setting of exposure**
 - Spread out over time and space (typical PulseNet identified)
- **Assess exposures of cases (ill) and controls (non ill)**
- **Identify which exposures are associated with disease**

Terms for graphic representation of analytical studies

E+

Exposed

E-

Non-exposed

D+

Diseased

D-

Non-diseased

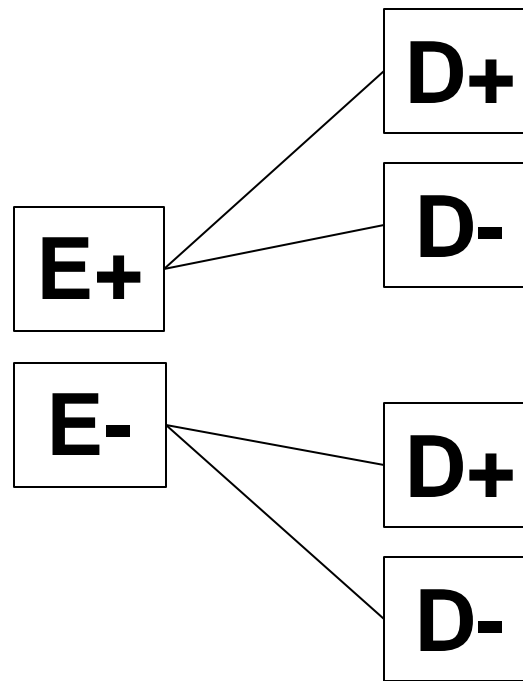


Study onset (selection of study population)



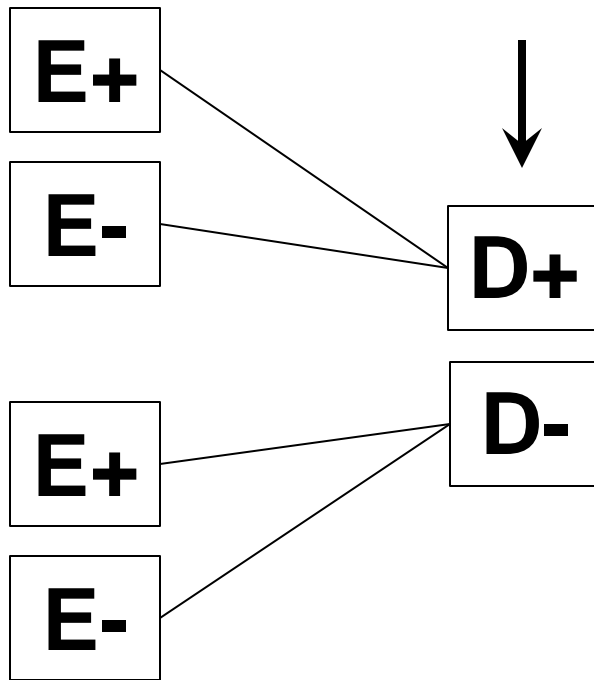
Time

Cohort study (retrospective)



Time

Case-control study



Time

What/who are controls and what is matching?

- **Controls**: subjects with whom comparison is made in a case control study
 - Selection is crucial to validity of study (but not quite as complicated as it seems)
- **Matching**: the process of making a study group and a comparison group comparable with respect to extraneous factors
 - Identifying controls based on a characteristic of a case
 - Age, location, medical condition
 - 1:1 matching or frequency-matching

The 2 x 2 Table

		Disease		
		Yes	No	
Exposure	Yes	a	b	a+b
	No	c	d	c+d
		a+c	b+d	n

- a. No. of subjects that are exposed and have the disease
- b. No. of subjects that are exposed and do not have the disease
- c. No. of subjects that are not exposed and have the disease
- d. No. of subjects that are not exposed and do not have the disease

Measures of association

A quantity that expresses the association between variables (i.e. exposure and disease)

- **Disease frequency**
 - Attack rate
- **Strength of association**
 - Relative risk (RR) used for cohort studies
 - Odds ratio (OR) used for case-control studies

Attack rate

- the proportion of subjects at risk that develop disease:
 - proportion of exposed subjects that become ill

$$= \frac{\text{\# ill among exposed}}{\text{total \# exposed}} \times 100$$

Attack rate example

"Petrification"

		Yes	No	
Eating Bertie Bott's Everyflavor Beans	Yes	23	37	60
	No	11	49	60
		44	86	120

Attack rate_{exposed}: $23 / 60 = 0.38 = 38\%$ (= Risk_{exposed})

Attack rate_{unexposed}: $11 / 60 = 0.18 = 18\%$ (= Risk_{unexposed})

Strength of association

- **Relative Risk (RR)**

The risk of getting sick if exposed to X
The risk of getting sick if not exposed to X

- **Odds ratio (OR)**

The odds that one was exposed to X, if ill
The odds that one was exposed to X, if not ill

- **OR is skewed**

- Larger than RR for same data

RR and OR formulas

		Disease		
		Yes	No	
Exposure	Yes	a	b	a+b
	No	c	d	c+d
		a+c	b+d	n

$$RR = \text{Risk}_{\text{exposed}} / \text{Risk}_{\text{unexposed}} = [a/a+b] / [c/c+d]$$

$$OR = \text{odds exposure}_{\text{diseased}} / \text{odds exposure}_{\text{non-diseased}} = (a:c) / (b:d) = \text{ad/bc}$$

RR and OR Example

"Petrification"

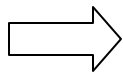
		Yes	No	
Eating Bertie Bott's Everyflavor Beans	Yes	23	37	60
	No	11	49	60
		34	86	120

$$RR = (23/60) / (11/60) = 2.1$$

$$OR = (23*49) / (37*11) = 2.8$$

Significance of association

- **P value**
 - probability that an association *at least as strong* as that observed might have arisen by chance alone
- **Confidence interval (CI)**
 - range within which 95% of times the true value of the estimated association lies (i.e. 95% CI)



Statistical significance depends on

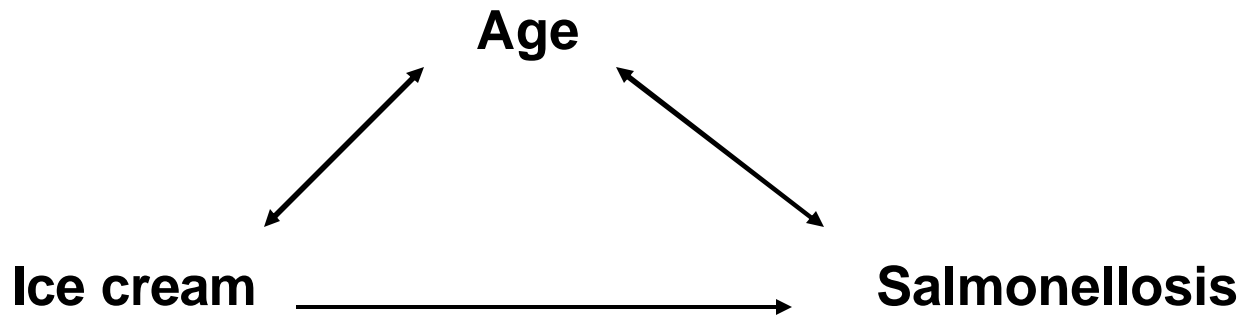
the difference between the groups

&

the number of people in the study

Confounding

- **Confounding** is the distortion of the apparent effect of an exposure brought about by the association with other factors that can influence the outcome
- **Confounders** are associated with both the exposure and the disease



Am I Biased?

- **Bias is error in measurement, not a predisposition to favor a particular study outcome**
 - **Recall bias**

“Who remembers what they ate yesterday let alone 3 weeks ago?”
 - **Information bias**

A misdiagnosed case or control

Stratified analysis

- Stratified analysis examines the association between an exposure and disease for different levels of another exposure

Chocolate Frog eaters:

Ate
Bertie Bott's
Beans

“Petrification”

	Y	N
Y	10	10
N	1	24

OR = 24.0

Chocolate Frog non-eaters:

Ate
Bertie Bott's
Beans

	Y	N
Y	13	27
N	10	25

OR = 1.2

Conclusions

- **Hopefully, you have a better understanding of the language of epidemiology**
- **Consumption of chocolate frogs (not Lord Voldemort) was the likely vehicle of transmission of “petrification” at Hogwart’s**

Thank you

Attack rate formula

		Disease		
		Yes	No	
Exposure	Yes	a	b	a+b
	No	c	d	c+d
		a+c	b+d	n

Attack rate_{exposed}: (diseased|exposed)/exposed = $a/(a+b)$

Attack rate_{unexposed}: (diseased|unexposed)/unexposed = $c/(c+d)$

P-value formula

		Disease		
		Yes	No	
Exposure	Yes	a	b	a+b
	No	c	d	c+d
		a+c	b+d	n

Chi-squared:
$$\frac{n(|ad-bc| - n/2)^2}{(a+b)(c+d)(a+c)(b+d)}$$

P-value = Percentiles of Chi-square distribution

P-value example

Salmonellosis

		Salmonellosis		
		Yes	No	
Eating product X	Yes	23	37	60
	No	11	49	60
		34	86	120

Chi-squared: 5.9

P-value < 0.05

Confidence Interval

		Disease		
		Yes	No	
Exposure	Yes	a	b	a+b
	No	c	d	c+d
		a+c	b+d	n

$$\text{Lower border} = \ln(\text{OR}) - Z^*(\text{var } \ln(\text{OR}))^{1/2} = L$$

$$\text{Upper border} = \ln(\text{OR}) + Z^*(\text{var } \ln(\text{OR}))^{1/2} = U$$

95% Confidence Interval for OR: e^L to $e^U = X$ to Y

Confidence interval example

Salmonellosis

		Salmonellosis		
		Yes	No	
Eating product X	Yes	23	37	60
	No	11	49	60
		34	86	120

$$\text{Lower border} = 1.03 - 1.96 * (0.18)^{1/2} = 0.2$$

$$\text{Upper border} = 1.03 + 1.96 * (0.18)^{1/2} = 1.9$$

95% Confidence Interval for OR: $e^{0.2}$ to $e^{1.9} = 1.2$ to 6.4

Interpretation RR and CI

- **RR > 1 : risk of disease among exposed greater than among unexposed**
- **RR = 1 : risk of disease the same among exposed and unexposed**
- **RR < 1 : risk of disease among exposed less than among unexposed (i.e. protective)**
- **$CI_{lower} < 1 < CI_{upper}$: no significant association**

Interpretation OR and CI

- **OR > 1 : odds of exposure among diseased greater than among non diseased**
- **OR = 1 : odds of exposure the same among diseased and non diseased**
- **OR < 1 : odds of exposure among diseased less than among non diseased (i.e. protective)**
- **$CI_{lower} < 1 < CI_{upper}$: no significant association**

Multivariate analysis: logistic regression, etc.

- **Fancy stratified analysis where you can control for many factors at once**
 - **Useful in foodborne studies where cases eat similar items**
 - **May be limited by number of people in study**