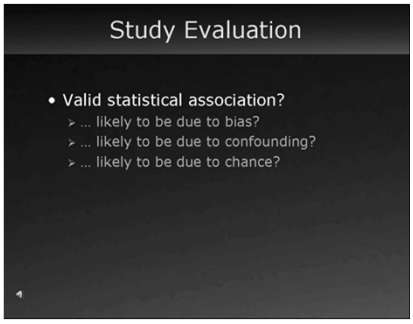
Interpretation: Bias

Learning Objectives

- Understand the aspects of a study to assess & to evaluate the validity of the study results

from: Module 7 Epi Assessment of Causation (slide 7)



Validity

- Evaluate the result
  - True?
  - False - alternative explanation

Validity

- What is it?
  - Validity = how close to TRUTH?
- Validity vs Reliability
  - Reliability = repeatability

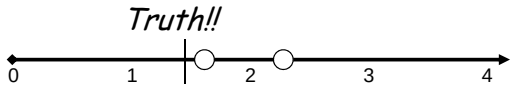
Validity

- If valid
  - Evaluate potential causal relationship
  - Generalizability

Validity

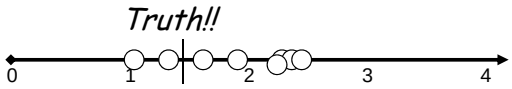
- Internal validity
  - Lack of measurement error
  - The association is correctly measured within the study
- External validity - generalizability

Validity



Measure of Association

Reliability



Measure of Association

"It is better to be vaguely correct than precisely wrong." – G. Frisvold

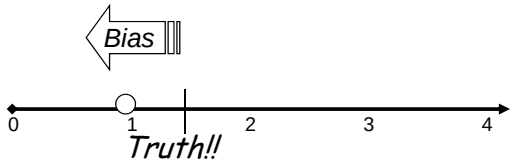
Validity

- Alternate explanations
  - Bias
    - systematic error in design or conduct leading to error in estimation of association
  - Confounding
    - A third variable distorting the association
  - Random Error
    - Not a systematic error, but chance or the luck of the draw

Bias

- A systematic error -> incorrect estimate of the measure of association
  - Create spurious association when there really is none (bias away from the null) or
  - Mask an association when there really is one (bias towards the null)
- Primarily introduced by investigator or study participants

Bias



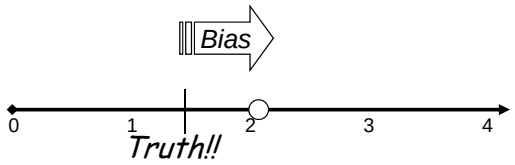
Measure of Association

Bias

- "Any systematic error in the
  - design,
  - conduct or
  - analysis
- of a study that results in a
  - mistaken estimate of the exposure's effect on the risk of disease."

- Schlesselman and Stolley, 1982

Bias



Measure of Association

Bias

- Bias ≠ prejudice
- All study types liable to bias
  - experimental, cohort, case-control
- Bias occurs in design and conduct of a study
- Can be evaluated but not fixed in the analysis

Biases

- Selection
- Misclassification
- Surveillance
- Observation
- Recall
- Interviewer
- Surrogate Interviews
- Loss to follow-up
- Non-response
- Temporal
- Analytic

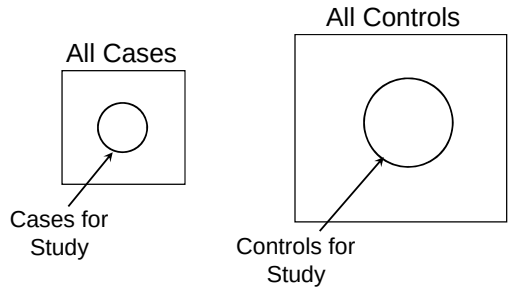
Selection Bias

- When:
  - Procedures to select cases and/or controls in a case control study or
  - the exposed and unexposed in a cohort study
  - biases the estimated RR or OR
    - The RR or OR is not what you would find if you looked at the whole population

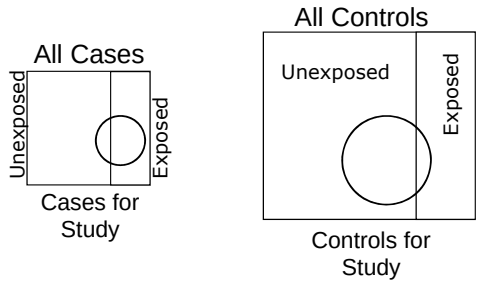
Selection Bias

- Chance that a case or a control will be selected depends on exposure status
  - Inclusion is not independent of exposure
- More likely to occur in case-control or retrospective cohort
  - exposure and outcome have already occurred at time of study selection

Case-Control Study-Selection Bias



Case-Control Study-Selection Bias



Coffee & Pancreatic Cancer

Case Control Selection Bias

- Example:
- Case control study:
  - Do PAP smears prevent cervical cancer?
  - Cases diagnosed at a city hospital
  - Controls randomly sampled from households same city
    - canvassing the neighborhood on foot

Case Control Selection Bias

	Cases	Controls
Pap Smear	100	150
No Pap Smear	150	100
	250	250

OR = (100)(100) / (150)(150) = .44

Interpretation:
In this study, women with cervical cancer were 56% less likely to have had a Pap smear than women in the neighborhood who did not have cervical cancer.

Case-Control Selection Bias

- Cases
  - from hospital
- Controls
  - neighborhood around hospital
- Bias
  - Only controls who were *at home* were included
  - Women at home
    - less likely to work
    - less likely to have regular checkups and PAP smears
  - Inclusion as a control was not independent of exposure

Case-Control Selection Bias

Selecting controls in the evening:

	Cases	Controls
Pap Smear	100	100
No Pap Smear	150	150
	250	250

OR = (100)(150) / (150)(100) = 1.0

No association of PAP smears and cervical cancer

Case-Control Selection Bias

- Women at home during the day
  - Not representative of the whole study population that produced the cases

They did not accurately represent the distribution of exposure in the study population that produced the cases

This led -> biased estimate of the association

Cohort Selection Bias

- Selection bias in cohort study
  - selection of exposed & unexposed subjects is not independent of outcome
  - so, it can only occur in a retrospective cohort study

Cohort Study Selection Bias

- Retrospective study
  - occupational exposure and a disease in a factory
- Exposed / unexposed enrolled on the basis of prior employment records

Cohort Study Selection Bias

- Records are old
  - many are lost
  - complete cohort not available for study
- If
  - people with exposure but not disease were more likely to have their records lost,
  - the estimate of association between the exposure and the disease will be inflated

	Dis +	Dis -
Exposed	a	b
Unexposed	c	d

Cohort Study Selection Bias

True relationship, if all records were available

	Disease	No Disease	
Exposed	50	950	1000
Unexposed	50	950	1000

RR = (50/1000) / (50/1000) = 1.00

Cohort Study Selection Bias

300 records lost, all among exposed, no disease

	Disease	No Disease	
Exposed	50	650	700
Unexposed	50	950	1000

RR = (50/700) / (50/1000) = 1.43

Selection Bias Solutions

- Once this bias has occurred:
  - Little or nothing can be done to fix it
- To avoid selection bias
  - Choose your population very carefully
    - Use same criteria for selecting cases and controls
    - obtaining all relevant subject records
    - obtaining high participation rates
    - account for diagnostic and referral patterns of disease
  - Secondary data to check validity

Selection vs Selection Bias

- Study Population:
  - Female,
  - Left-handed,
  - Epidemiologists
- Selected?
  - yes, very
- Selection ***Bias***
  - not necessarily!
- but, *Generalizability*?

Selection Bias Summary

- Selection Bias
  - Cases/controls or exposed/unexposed chosen so that
    - there appears to be a relationship of exposure and outcome that doesn't really exist or
    - a relationship that does exist is masked

Biases

- ✓ Selection
- Misclassification
- Surveillance
- Observation
- Recall
- Interviewer
- Surrogate Interviews
- Loss to follow-up
- Non-response
- Temporal
- Analytic

Misclassification

- Errors in ascertainment of exposure or outcome

		Disease Status		
		Yes	No	TOTAL
Exposure Status	Yes	a	b	$a+b$
	No	c	d	$c+d$
	TOTAL	$a+c$	$b+d$	$a+b+c+d$

Misclassification

- Nondifferential or random
  - Non-selective or random error
  - Effect is the same in both groups

Decreases likelihood of finding an association if there is truly an association

Misclassification

- Differential
  - Likelihood of error in exposure classification
    - different for those with or without the outcome

  OR
  - Likelihood of error in outcome classification
    - different for those who are / are not exposed

  OR
  - Likelihood of error for one variable varies according to the actual value of the other
- Effect
  - Toward the null
  - Away from the null – in either direction
  - unknown

Misclassification Example 1

Truth

	Disease	No Disease	
Exposed	200	100	
Unexposed	100	200	
	300	300	

OR = (200)(200) / (100)(100) = 2.0

Misclassification Example 1

Differential Misclassification

	Disease	No Disease	
Exposed	200	50	
Unexposed	100	250	50
	300	300	

OR = (200)(250) / (100)(50) =10

True OR = 2

Misclassification Example 2

Truth

	Disease	No Disease	
Exposed	200	100	
Unexposed	100	200	
	300	300	

OR = (200)(200) / (100)(100) = 4.0

Misclassification Examples

Non-Differential Misclassification

10% of cases & 10% of controls misclassified

	Disease	No Disease	
Exposed	180	90	10%
Unexposed	120	210	
	300	300	

10%

OR = (180)(210) / (90)(120) = 3.5

True OR = 4

Misclassification Solution

- Techniques to verify status:
  - Narrow case definitions
    - Clinical / histologic verification
    - Validation sub-studies
  - Refined methods of exposure measurement
    - Biologic measures
    - Lab techniques

Surveillance Bias

- Outcome often sub-clinical or asymptomatic
  - chances of diagnosis greater in those who are seen more often

Surveillance Bias Solutions

- Stratify
  - Create an index of medical care usage
  - Run separate analyses for each level of use
- Prospective study
  - Systematically assess for outcome in all participants

Stratified Analysis

Low Use

		Disease Status: Yes	No	TOTAL
Exposure Status	Yes	*a*	*b*	*a + b*
	No	*c*	*d*	*c + d*
	TOTAL	*a + c*	*b + d*	*a + b + c + d*

High Use

		Yes	No	TOTAL
Exposure Status	Yes	*a*	*b*	*a + b*
	No	*c*	*d*	*c + d*
	TOTAL	*a + c*	*b + d*	*a + b + c + d*

Observation Bias

- Systematic differences in the WAY information on exposure or disease is obtained
- -> incorrect classification
  - exposed / unexposed or
  - diseased / not diseased
- Occurs after the subjects have entered the study

Recall Bias

TRUTH

	Case	Control
Ex.	40	20
Unex.	60	80
	100	100

Odds Ratio: 2.7

OBSERVED

	Case	Control	
Ex.	40	10	10
Unex.	60	90	
	100	100	

Odds Ratio: 6.0

Recall Bias - Solutions

- Sick controls
- Standardized questionnaires
  - that obtain complete information
- Mask subjects to study hypothesis

Interviewer Bias

- Systematic difference in
  - soliciting,
  - recording,
  - interpreting information
    - Exposure information sought when outcome is known
      Case control
    - Outcome information sought when exposure is known
      Cohort study

Interviewer Bias Solutions

- Standardize interviewer training
- Standardize interview procedure
- Monitor quality during data collection (taping)
- Mask interviewer
  - to study hypothesis
  - disease / exposure status of subjects

Surrogate Interviews

- Cases unavailable for interview
  - Family member interviews
    - Does family member know about exposure?
    - Occupational exposure?
- Controls available for interview

Surrogate Interviews - Solutions

- Standardize data collection for case / control groups
  - Family member interviews for all
  - Compare family member interview to control interview

Loss to Follow up

- People who are lost to follow up may differ from those that remain in the study
  - If subjects lost differ in outcome and exposure from those that remain -> BIAS
- Solution
  - Don't lose people!
  - Achieve high and equal rates of follow up for the exposed and unexposed groups

Non-Response Bias

- Responders $\overset{?}{=}$ Non-responders?
- Evaluation
  - Compare responders to non-responders for each variable on which you have data

Temporal Bias

- "Exposure"= early manifestation of disease
- Solution: stratify by early manifestation

Analytic Bias

- Wrong choice of statistical techniques
  - Example: data don't meet assumptions of test
- Solution
  - Learn more statistics yourself OR
  - Consult a biostatistician
    - Best done in DESIGN Phase
    - Even the best statistician cannot help you if you have not collected the appropriate data in the right way….

Issues in Interpretation

this class
- Could bias have occurred?
  - Consider the design and conduct of study

this class
- Is bias actually present?
  - Evaluate where possible
    - Compare those lost with those kept

beyond this class
- Consequences of bias large enough to really alter the measure of association?
  - Sensitivity analysis

beyond this class
- Which direction is the distortion?
  - Towards or away from the null?

Potential for Bias

Probability of:	Ecological	Cross Sectional	Case-Control	Cohort
Selection Bias				
Recall Bias				
Loss to follow-up				
Confounding				

variable none less more MOST

Interpretation: Confounding

Introduction

60

Learning Objectives

- Understand the concepts of confounding and interaction, how to detect them and distinguish them from each other

61

Confounding

- Definition
  - mixing of the effect of the exposure on the disease with that of a third factor
    - Thus
      - It is an alternate explanation for an observed association between an exposure and disease
    - Result
      - distorts the true association toward the null (negative confounding) or away from the null (positive confounding).

62

Confounding Example

	Florida	Alaska	RR
Number of Deaths	131,044	2,064	
Total Population	12,335,000	524,000	
Crude Mortality Rate (per 100,000)	1,062.4	393.4	2.7
Age Adjusted Mortality Rate (per 100,000)	812.4	764.8	1.1

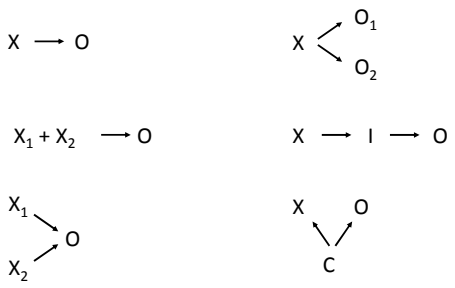
63

Confounder Requirements

- Is related to exposure
- Is related to disease
- Is not a biologic consequence of the exposure
- Is not part of the causal pathway

64

Possible Associations



X = exposure, O = outcome,
C = confounder, I = Intermediate

65

MI and Physical Activity

- Could age be a potential confounder?
  - Is age related to myocardial infarction (MI)?
  - Is age related to PA?
- Could fluid intake be a potential confounder?
  - Is fluid intake related to MI?
  - Is fluid intake related to PA?

66

Confounding Notes

- Association of potentially confounding factor (PCF) & disease
  - does not have to be a causal
- PCF must be predictive of disease, independent of its association with the exposure under study
  - must be association of PCF and D among the non-exposed

67

Why worry about confounding?

- Confounding factors = nuisance variables
  - Distort the relationship you want to measure
  - Need to remove their effect
    - Age-standardization is an example

68

Causal Pathway

- A variable cannot be a confounder if it is a step in the causal chain or pathway
- Example
  - Moderate alcohol consumption increases serum HDL levels
  - Elevated serum HDL levels decrease risk of heart disease
    - HDL level is a step in this causal chain – an Intermediary
    - not a confounder
  - We don't need to control for it
  - It is something interesting that helps us understand the disease mechanism

69

Interpretation: Confounding

Strategies in Design

70

Learning Objectives

- Understand how to avoid confounding in the design of a study and
- Correct for confounding in the analysis

71

Avoiding Confounding

- Must be considered during design AND analysis
- Must know something about the potential confounding factors before you design or analyze

72

Design Strategies

- Selection or restriction
- Randomization
- Matching
  - Individual matching or frequency matching

73

Selection/Restriction

- Include only one strata in the study
  - If gender was the confounding variable
    - Include only one gender, e.g. men only
- Advantages?
- Disadvantages?

74

Randomization

- Typically in clinical trials
  - Randomize the exposure
  - Generally leads to an even distribution of subjects with confounding characteristics between treatment groups

75

Matching

- Choice of study subjects
  - potential confounders distributed in an identical manner among
    - exposed and unexposed groups (cohort study)
    - cases and controls (case control study)

76

Matching Example

- Cohort study
  - Exercise and heart attack
    - exercisers and non-exercisers
    - Confounders
      - age, sex, smoking
  - Exposed subject
    - 45 year old nonsmoking female
    - Choose unexposed subject
      - Age 45 (45 + or – a couple of years)
      - female
      - Non smoker

77

Matching

- Total sample size required usually smaller than unmatched design
- Does not control for other confounders

78

When to Match?

- Insufficient subjects for control through stratification/multivariate analysis
- May also control for factors that are indefinable and difficult to quantify
- Feasibility for accruing control group
- Small case series

79

Confounding in the Analysis

Is eating pizza related to MI?

		Myocardial Infarction		
		Yes	No	TOTAL
Eats Pizza	Yes	91	19	110
	No	19	91	110
	TOTAL	110	110	220

Crude Odds ratio = $\frac{91*91}{19*19} = 23$

80

Confounding in the Analysis

- Is this a valid result?
- Could there be another explanation for this large effect?
- Could drinking beer be a potential confounder and be an explanation for the effect?

81

Detecting Confounding

- Is the PCF related to both exposure and disease?
- Is the crude relationship between our exposure of interest different than the relationship after adjusting for the potential confounder?

82

Confounding Criterion 1

Is the PCF related to both exposure and disease?

Is Beer drinking related to eating pizza (exposure)?

	Beer	No Beer
Pizza	95	15
No Pizza	15	95
TOTAL	110	110

Odds ratio = 40.1

83

Confounding Criterion 1

Is the PCF related to both exposure and disease?

Is Beer drinking related to MI (outcome)?

	Beer	No Beer
MI	80	30
No MI	30	80
TOTAL	110	110

Odds ratio = 7.1

Could Beer Drinking be in the pathway?

84

Confounding Criterion 2

Is the crude association similar to the association within the strata of the PCF?

Beer Drinkers

	MI	No MI
Pizza	75	20
No Pizza	5	10
TOTAL	80	30

Odds ratio = 7.5

Beer Non-drinkers

	MI	No MI
Pizza	10	5
No Pizza	20	75
TOTAL	30	80

Odds ratio = 7.5

Odds Ratio$_{crude}$ = 23

85

Note

- All adjustment formulas and procedures assume that the strata-specific effects are equal

86

Interpretation: Confounding

Strategies in Analysis

87

Learning Objectives

- Understand how to avoid confounding in the design of a study and
- Correct for confounding in the analysis

88

Analytic Strategies

- Matching
  - *Required* if matching used in design
- Stratification
  - Suitable for all designs
  - To detect & control confounding AND to detect & describe effect modification
- Multivariate analyses

89

Matched Analysis

		Control: Exposed	Control: Unexposed
Case: Exposed		a	b
Case: Unexposed		c	d

- Compare discordant pairs
  - OR = B/C
  - McNemar's odds ratio

90

Stratification

- Evaluation of the exposure / outcome association within homogeneous categories (strata) of the confounding variable

91

Example

- Case Control Study
  - Exposure: Oral contraceptive use
  - Outcome: Heart attack

Total Data in one 2 x 2 table)

		Case	Control
OC Use	Yes	39	24
	No	114	154

Crude OR = 2.2

92

Stratification

		Age< 40		Age 40 and over	
		Case	Control	Case	Control
OC Use	Yes	21	17	18	7
	No	26	59	88	95
		Stratum-specific OR = 2.8		Stratum-specific OR = 2.8	

- Each stratum like a restricted analysis
- Stratum specific ORs do NOT equal the crude
  - Stratum specific = 2.8, Crude = 2.2
- Difference > 10%

93

Adjustment

- CONTROL by computing an adjusted (summary) effect measure
  - Mantel-Haenszel Adjusted Odds Ratio

$$OR_{MH} = \frac{\sum (a_i \times d_i / N_i)}{\sum (b_i \times c_i / N_i)}$$

  - Logistic Regression

94

OR_{MH}

Beer Drinkers			Non Beer drinkers		
	MI	No MI		MI	No MI
Pizza	75	20	Pizza	10	5
No Pizza	5	10	No Pizza	20	75
TOTAL	80	30	TOTAL	30	80
	Odds ratio = 7.5			Odds ratio = 7.5	

Odds $Ratio_{crude}$ = 23

$$OR_{MH} = \frac{\sum (a_i \times d_i / N_i)}{\sum (b_i \times c_i / N_i)} = \frac{(75*10)110 + (10*75)110}{(5*20)110 + (20*5)110}$$

$$= 7.5$$

95

Beer Drinking is a Confounder

- Beer drinking is related to both exposure (pizza) and disease
- Crude OR does not equal strata-specific OR
  - Crude OR = 23
  - OR for beer drinkers = 7.5
  - OR for non-beer drinkers= 7.5
  - OR_{MH} = 7.5
- Adjusted OR does not equal crude OR, but does equal strata-specific OR

96

Defining "Different"

- When you, the investigator, think the difference is meaningful
- A suggested general rule: when the difference between ORs >10%

97

Is this evidence of confounding?

1. Crude OR = 2.1
   - Strata OR1 approximately = 1.5
   - Strata OR2 approximately = 1.8
   - Adjusted OR =
2. Crude OR=2.1
   - Strata OR1 approximately = 2.3
   - Strata OR2 approximately = 1.9
   - Adjusted OR=

98

Limits of Stratification

- It is difficult to control for many variables simultaneously because a large number of strata will be generated relative to the number of study subjects

99

Example

- Case-control study of physical activity and heart disease
  - Stratify by
    - Gender
      - two categories: male and female
    - Age
      - five categories: 40-49, 50-59, 60-69, 70-79, 80 and over,
    - Smoking status
      - three categories: never smoked, light smoker, heavy smoker
  - you will end up with 30 strata.
- If you have a study with a few hundred cases and controls, you will end up with small numbers or even zeros in many cells

100

Multivariate Analysis

- Statistical technique
  - simultaneously adjusts for several variables
- Construction of mathematical model
  - describes associations between exposure, disease, and confounders
- Examples
  - multiple linear regression for continuous variables
  - logistic regression for case-control data
  - Cox proportional hazards model for cohort data

101

Multivariate Approach

- Logistic regression

$$\ln({}^{Y}/_{1-Y}) = \alpha + \beta_1 x_1 + \ldots + \beta_n x_n + \varepsilon$$

$$\text{OR}(x_1) = e^{\beta_1}$$

102

Summary of Confounding

- Confounding can either exaggerate or minimize the true association
- Epidemiologists have developed many methods to control confounding in the design and analysis

103

Summary of Confounding

- Mixing of effect between and exposure, an outcome, and a third variable known as a confounder
- Considered a nuisance
- Studies may have a small, moderate or large degree of confounding

104

Remember this earlier note?

- All adjustment formulas and procedures assume that the strata-specific effects are equal
- What happens if the strata-specific effects are NOT equal?

105

Interpretation: Interaction

Learning Objectives

- Understand interaction & how to distinguish this from confounding

Effect Modification

- Interaction = effect modification
  - the ***magnitude*** of the association between the exposure of interest and the outcome is ***different*** for each level of the third variable
  - If strata-specific effects are not approximately equal, then effect modification is present

$Strata_1 \neq ... \neq Strata_n$

OR or RR in women = 5.2
OR or RR in men = 1.1

Effect Modification = Interaction

Detecting Effect Modification

- Compare crude estimates with strata-specific estimates
  - If strata-specific estimates are approximately equal, then no effect modification.
    - Evaluate for confounding and adjust if appropriate
  - If strata-specific estimates are not approximately equal, then there is effect modification
    - No need to evaluate for confounding
    - Describe

SBP / MI / Gender Example

Crude RR = (95/391) / (173/1240)
= 1.74 (1.39-2.18)

SBP	MI	No MI	
≥ 165	95	296	391
< 165	173	1067	1240

Men

SBP	MI	No MI	
≥ 165	44	106	150
< 165	120	537	657

Strata-specific RR = 1.61 (1.2-2.2)

Women

MI	No MI	
51	190	241
53	530	583

Strata-specific RR = 2.33 (1.6-3.3)

SBP / MI / Gender Example

- Is gender and effect modifier?
  - Strata-specific RR are not approximately equal, although they are both in same direction
    - RR for men = 1.6
    - RR women = 2.3

When is 'different' different?

- Epidemiological perspective
  - When you, the investigator, judge the difference to be of importance
- Statistical perspective
  - When the difference is statistically significant

Presenting Results

- If there is effect modification present
  - Do NOT summarize (adjust) the association
  - Describe the associations
- Effect modification usually tells something about the causes

SBP / MI / Gender Example

- Crude RR = 1.7
  - Male RR = 1.6
  - Female RR = 2.3
- Adjusted RR = 1.9

Example

- Relation of
  - BMI
    - body mass index, a measure of obesity
  - and breast cancer
  - varies with
  - menopausal status
    - Pre-menopausal women:
      - higher BMI decreases risk
    - Post-menopausal women:
      - higher BMI increases (or does not affect) risk

Summary-1

- Why stratify?
  - To detect and describe effect modification
  - To detect and remove confounding
- When do you present strata-specific results?
  - When you have effect modification
  - RR_{crude} ne $RR_{strata1}$ ne $RR_{strata2}$

ne = not equal to, eq = equal to

Summary-2

- When do you present adjusted rates?
  - When you have confounding and no effect modification.
  - RR_{crude} **ne** $RR_{strata1}$ **eq** $RR_{strata2}$
- When do you present the crude results?
  - When there is no confounding and no effect modification

ne = not equal to; eq = equal to

Summary -3

- All strategies still require the investigator to know in advance which potential confounding factors or effect modifiers need to be included in data collection
  - You can't adjust for variables that you did not collect data on….

Confounding / Interaction Step by Step

1. Is it in the causal pathway?
   - Yes STOP! This is neither confounding nor interaction
   - No continue to #2
2. Is it associated with the exposure of interest?
   - No STOP! This is neither confounding nor interaction
   - Yes continue to #3
3. Is it associated with the outcome?
   - No STOP! This is neither confounding nor interaction
   - Yes continue to #4

Confounding / Interaction Step by Step

4. Does it affect the measure of association between the exposure of interest and the outcome (i.e., is $OR_{crude} \neq OR_{strata}$)?
   - No STOP! This is neither confounding nor interaction
   - Yes – continue to #5
5. Are the strata specific OR/RR approximately equal?
   - Yes – you have confounding, present an adjusted measure of association
   - No – you have INTERACTION! Present strata specific measures of association separately – this has become the main point of your publication!

Confounding / Interaction Step by Step

What your ORs or RRs look like:	Effect Modification	Confounding	Neither	You Should Present:
Crude =Strata$_1$ = … = Strata$_n$			X	Crude
Crude ≠ Strata$_1$ = … = Strata$_n$		X		Summary (Mantel-Haenszel)
Crude ≠ Strata$_1$ ≠ Strata$_n$	X			Strata specific - do not present a summary

Standardization

Definition & Motivation

122

Learning Objectives

- Understand motivation for standardization

123

Disease Rates

- Rates are usually calculated because you want to COMPARE
  - Compare some observed experience with a target rate
  - Compare two populations at the same time period
  - Compare same population at two different time periods

124

Types of rates

- Crude rates
- Specific rates
  - Age, cause, time period
- Adjusted rates
  - Age-adjusted or age standardized rate

125

Crude rates

- Summary rates for whole population
- Denominators usually some midpoint estimation of population

126

Specific Rates

- Cause-specific
- Age-specific
- Race-specific
- Sex-specific

- Numerators and denominators are restricted to population subsets

127

10 Leading Causes of Death

Ages 25-44, All races, Both Sexes, US, 1991

Rank Order	Cause of death	#	PMI*	Cause specific death rates**
1	Accidents /adverse events	26,596	18.0	32.2
2	Malignant neoplasms	22,228	15.0	27.0
3	HIV/infection	21,747	14.7	26.4
4	Heart disease	15,822	10.7	19.2
5	Homicide / legal intervention	12,372	8.4	15.0
6	Suicide	12,281	8.3	14.9
7	Chronic liver disease & cirrhosis	4,449	3.0	5.4
8	Cerebrovascular diseases	3,343	2.3	4.1
9	Diabetes mellitus	2,211	1.5	2.7
10	Pneumonia & influenza	2,203	1.5	2.7
	All causes	147,750		

Adapted from National Center for Health Statistics, 1991, Monthly Vital Statistics Report, 42(2):21. 1993

128

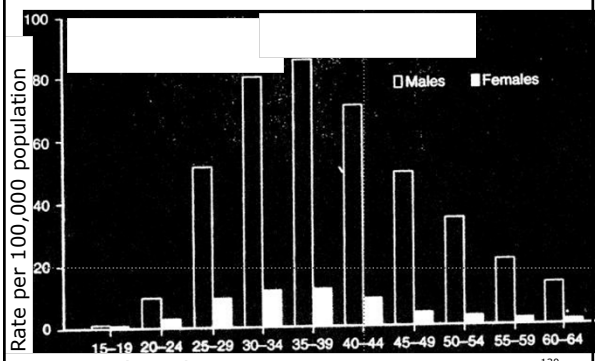
Cause-Specific Death Rate

Total number of deaths due to a specific cause among individuals ages 25-44 in the US in 1991

Total number of individuals in the US population in 1991 ages 25-44 years

129

AIDS Incidence by Age and Gender, US 1992



130

Table 3-4 15-Year Trends: Age-Specific Cancer Incidence Rates* (by Sex, Age Group, and Year of Diagnosis, SEER Program, 1973–1987; All Sites Combined, All Races)

	Year of Diagnosis														
Sex/Age	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987
Males and Females	319.9	332.3	331.8	337.2	336.6	336.7	340.2	344.0	349.7	350.2	355.4	362.4	368.9	369.9	377.2
0–54:	97.2	101.7	98.7	100.7	99.9	98.9	98.0	97.5	98.3	97.3	97.8	100.9	103.7	105.4	106.6
0–14	12.6	13.0	11.3	12.7	12.7	13.0	12.9	12.7	12.1	12.8	12.7	13.6	14.0	13.7	13.4
15–34	35.8	37.8	38.1	37.5	39.0	38.2	37.9	37.9	38.2	38.5	39.1	39.4	40.7	41.7	40.0
35–44	160.3	159.4	159.2	151.8	156.5	159.3	152.1	154.3	155.5	156.6	160.8	169.3	172.2	176.1	178.0
45–54	405.0	431.8	413.9	433.3	419.3	410.9	412.7	407.6	412.5	402.1	399.9	410.5	422.7	429.8	441.2
55–64	844.4	896.7	900.6	899.6	887.9	896.9	893.4	897.5	925.3	929.7	943.0	955.1	982.5	953.4	961.7
65+:	1660	1700	1717	1756	1767	1788	1814	1853	1878	1887	1924	1968	1976	1998	2055
65–74	1457	1483	1496	1520	1532	1529	1583	1614	1632	1650	1686	1720	1741	1773	1806
75+	1992	2054	2076	2139	2151	2158	2192	2242	2280	2274	2312	2345	2359	2364	2461
Males	365.2	372.1	377.8	388.9	392.7	394.8	400.5	407.5	411.5	411.7	419.2	423.0	428.8	430.7	441.9
0–54:	77.5	79.4	77.9	80.4	83.0	81.6	81.4	83.2	82.0	81.5	83.3	84.9	87.7	90.9	91.1
0–14	14.1	14.3	12.3	13.6	12.7	13.9	13.3	12.7	12.9	13.6	13.8	13.9	14.9	15.0	14.1
15–34	30.8	31.4	32.3	30.9	32.7	31.9	33.0	33.8	33.8	33.4	36.1	36.3	38.0	40.0	38.0
35–44	106.0	99.8	104.7	102.2	115.5	107.9	104.4	107.6	106.4	111.6	117.0	121.1	128.9	132.3	136.4
45–54	328.5	346.3	333.3	353.9	356.4	353.7	354.0	363.3	355.2	346.2	345.9	352.1	356.9	371.4	375.6
55–64	882.6	910.0	924.4	948.2	936.6	960.2	963.2	979.1	1010	1007	1024	1032	1059	1019	1025
65+:	2245	2274	2331	2401	2429	2439	2496	2538	2560	2566	2613	2632	2623	2672	2779
65–74	1854	1884	1914	1948	1991	1997	2040	2071	2102	2100	2142	2164	2165	2213	2283
75+	2883	2911	3012	3139	3143	3160	3239	3298	3305	3331	3381	3364	3370	3419	3588
Females	293.2	311.0	308.7	307.7	303.5	302.3	304.3	305.4	312.3	313.4	316.5	326.3	334.5	333.4	338.4
0–54:	116.1	123.1	118.6	120.2	116.2	115.5	114.0	111.4	114.0	112.8	111.8	116.6	119.1	119.5	121.7
0–14	11.0	11.8	10.3	11.9	12.6	12.0	12.5	12.7	11.3	12.0	11.5	13.4	13.0	12.3	12.7
15–34	40.8	44.1	43.8	44.1	45.3	44.4	42.6	41.9	42.5	43.5	42.1	42.8	43.4	43.3	42.0
35–44	212.7	217.2	212.2	200.1	196.3	209.4	198.8	200.1	203.7	200.8	203.7	216.7	214.8	219.1	218.9
45–54	477.4	513.1	490.4	508.8	479.2	465.3	468.5	449.7	467.1	455.6	451.5	466.4	485.7	485.5	504.3
55–64	809.5	865.1	879.3	856.0	844.1	839.8	830.6	824.2	849.8	860.4	870.4	885.8	914.5	894.8	905.0
65+:	1268	1320	1318	1337	1339	1336	1377	1416	1441	1453	1483	1528	1563	1568	1591
65–74	1148	1173	1176	1192	1181	1170	1231	1262	1267	1300	1330	1375	1410	1429	1432
75+	1463	1559	1550	1574	1597	1607	1617	1668	1724	1702	1732	1777	1813	1793	1849

*Rates are per 100,00 and are age adjusted to the 1970 US standard population. Each rate has been age-adjusted by 5-year age group.
Source: Reprinted from CL Kosary et al, *SEER Cancer Statistics Review, 1973–1992: Tables and Graphs*, p II.23, 1995, NIH Pub No 96-2789, National Cancer Institute.

131

Crude Death Rate / Life Expectancy

County (year)	Crude Death Rate (per 1,000)	Life Expectancy @ birth (years)
Colombia (1985 – 1989)	7.4	63.4
US (1988)	8.8	71.3
Swaeden (1989)	10.8	74.2

United Nations. Demographic Yearbook. New York, United Nations, 1990.

132

Comparing Two Rates

- Category specific comparisons
- Comparisons with adjusted rates
  - Direct adjustment
  - Indirect adjustment

133

Crude & Age Specific Mortality Rates

Age (years)	# Cancer Deaths	Population July 1, 1980	Mortality (per 100,000)
Under 5	686	16,548,000	4.2
5 - 9	777	16,700,000	4.7
10 - 14	720	18,242,000	3.9
15 - 19	1145	21,168,000	5.4
20 - 24	1638	21,319,000	7.7
25 - 29	2041	19,521,000	10.5
30 - 34	3040	17,561,000	17.3
35 - 39	4684	13,965,000	33.5
40 - 44	7786	11,669,000	66.7
45 - 49	14,230	11,090,000	126.5
50 - 54	26,300	11,710,000	221.9
55 - 59	41,600	11,615,000	338.2
60 - 64	53,045	10,083,000	525.6
65 - 74	127,430	15,281,000	817.9
75+	130,959	9,959,000	1313.4
Total (crude rate)	416,481	226,546,000	183.8

US Census Bureau, Statistical Abstract of the United States: 1984 (104th ed.); and USDHH Vital Statistics of the US, 1980 Vol. II Mortality, Part B

134

Crude Rates

- Are summary rates of the total population
- Weighted average of the population
  - Weighted by the population structure
- Sum of specific rates x population in the specific categories

135

Why Adjust?

- Want to compare rates
- Populations to compare differ in distribution on another variable(s) (e.g. age) that may have a relationship with the outcome
- Strata-specific comparisons tedious & don't provide a summary
- Standardization
  - accounts for these differences so that we can make valid comparisons

136

Rate Standardization

137

Comparing Rates

Has the cancer mortality rate changed from 1940 to 1980?

Crude Mortality Rates

- 1940
  - $120.2/10^5$/yr
- 1980
  - $183.8/10^5$/yr

138

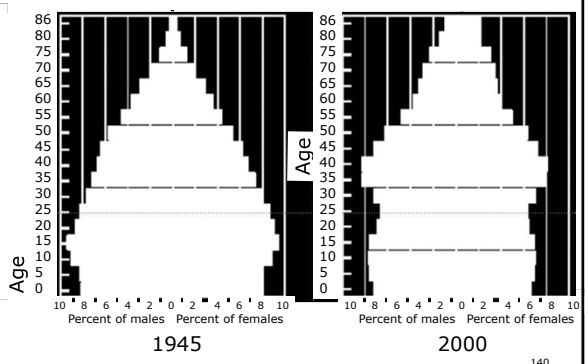
Crude & Age Specific Cancer Mortality Rates, US 1940 & 1980

Age (years)	Cancer mortality rates (per 100,000) 1940	1980
Under 5	4.7	4.2
5 – 9	3.0	4.7
10 – 14	2.9	3.9
15 – 19	4.0	5.4
20 – 24	6.8	7.2
25 – 29	11.6	10.5
30 – 34	23.5	17.3
35 – 39	43.4	33.5
40 – 44	80.3	66.7
45 – 49	133.4	126.5
50 – 54	209.0	228.9
55 – 59	309.9	338.2
60 – 64	443.3	525.6
65 – 74	695.1	817.9
75+	1183.5	1393.4
Total (crude rate)	120.2	183.8

USDHH *Vital Statistics of the US*, 1980 Vol. II Mortality, Part B

139

Population Pyramids



140

How to compare two rates

- Category specific comparisons
- Comparisons with adjusted rates
  - Direct adjustment
  - Indirect adjustment

141

Crude Rate

Strata	Population	Events
1	20	6
2	40	4
3	60	2
Total	120	12

$$\text{Rate per } n = \frac{\text{Events} * n}{\text{Population}}$$

$$\text{Rate per 1,000} = \frac{\quad}{\quad}$$

142

Strata Specific Rates

Strata	Population	Events	Rate (per 1,000)
1	20	6	300
2	40	4	100
3	60	2	33
Total	120	12	100

$$\text{Rate}_{\text{stratum1}} \text{ per } n = \frac{\text{Events}_{\text{stratum1}} * n}{\text{Population}_{\text{stratum1}}}$$

$$\text{Rate}_{\text{stratum1}} \text{ per } \quad = \frac{\quad}{\quad} = 300$$

143

Strata Specific Rates

Strata	Population	Events	Rate (per 1,000)
1	20	6	300
2	40	4	100
3	60	2	33
Total	120	12	100

$$\text{Rate}_{\text{stratum2}} \text{ per } n = \frac{\text{Events}_{\text{stratum2}} * n}{\text{Population}_{\text{stratum2}}}$$

144

Direct Adjustment

- Category-specific rates x standard population distribution
- Provides an estimate of what would be expected if both populations in the comparison had the same population distribution

145

Direct Adjustment

	Population A			Population B			Expected		
Strat	Pop	Events	Rate (per 1,000)	Pop	Events	Rate (per 1,000)	Strd Pop	A	B
1	20	6	300	60	2	33	2000	600	
2	40	4	100	20	2	100	4000		
3	60	2	33	30	3	100	1000		
Total	120	12	100	110	7	64	7000		

*

Expected Events$_{PopA,strat1}$ = Rate$_{PopA\ strat1}$ * Standard Pop$_{strat1}$

= $\frac{\text{Events}_{stratum1}}{\text{Population}_{stratum1}}$ * Standard Pop$_{stratum1}$

= ——— * =

146

Direct Adjustment

	Population A			Population B			Expected		
Strat	Pop	Events	Rate (per 1,000)	Pop	Events	Rate (per 1,000)	Strd Pop	A	B
1	20	6	300	60	2	33	2000	600	67
2	40	4	100	20	2	100	4000	400	400
3	60	2	33	30	3	100	1000	33	100
Total	120	12	100	110	7	64	7000	1033	567

Adjusted Rate$_{Pop\ A}$ = $\frac{*\sum_1^3 \text{Expected Event}_{stratum}}{\text{Total Standard Population}}$ = $\frac{1033}{7000}$

= 147.6 per 1,000

147

**Crude, handmade summation sign – fancy symbols don't always come across!*

Direct Adjustment

	Population A			Population B			Expected		
Strat	Pop	Events	Rate (per 1,000)	Pop	Events	Rate (per 1,000)	Strd Pop	A	B
1	20	6	300	60	2	33	2000	600	67
2	40	4	100	20	2	100	4000	400	400
3	60	2	33	30	3	100	1000	33	100
Total	120	12	100	110	7	64	7000	1033	567

Adjusted Rate$_{Pop\ B}$ = $\frac{*\sum_1^3 \text{Expected Event}_{stratum}}{\text{Total Standard Population}}$ = $\frac{567}{7000}$

= 81.0 per 1,000

148

**Crude, handmade summation sign – fancy symbols don't always come across!*

Comparing Rates

Has the cancer mortality rate changed from 1940 to 1980?

Crude Mortality Rates

- 1940
  - 120.2/10^5/yr
- 1980
  - 183.8/10^5/yr

149

Age-Adjustment

Calculation of age-adjusted cancer mortality rates in the US using the 1940 US population as the standard

Age Group	1940 Population (1,000's)	Cancer Mortality Rates, 1980	Expected Deaths in 1980
Under 5	10,541	4.2	442.7
5 - 9	10,685	4.7	502.2
10 - 14	11,746	3.9	458.1
15 - 19	12,334	5.4	656.0
20 - 24	11,588	7.3	864.3
25 - 29	11,097	10.5	1,165.2
30 - 34	10,242	17.3	1,771.2
35 - 39	9,548	33.5	3,197.6
40 - 44	1,700	66.7	5,861.6
45 - 49	1,256	126.5	10,891.2
50 - 54	7,257	298.9	16,611.3
55 - 59	5,344	338.2	20,956.2
60 - 64	4,798	525.6	24,889.8
65 - 74	6,377	817.9	52,157.5
75+	2,635	1,313.4	34,731.1
Total (crude rate)	131,670	183.8	174,775.7

Age-adjusted Cancer mortality rate

174,773.7
131,670,000

=132.7/10^5

150

Comparing Rates

Has the cancer mortality rate changed from 1940 to 1980?

- 1940
  - $120.2/10^5$/yr
- 1980
  - $183.8/10^5$/yr (crude)
  - $132.7/10^5$/yr (adjusted)

151

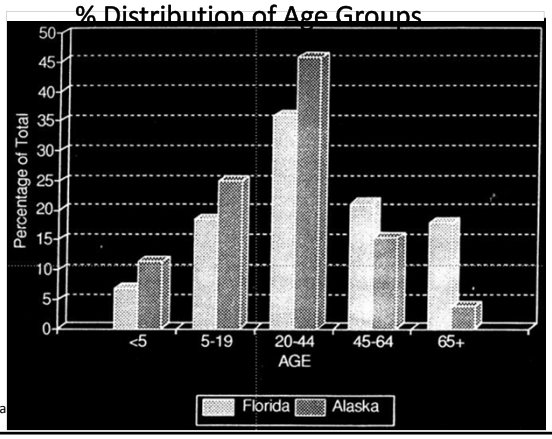
Comparing Rates

Are people more likely to die in Florida than in Alaska?

	Florida	Alaska
Number of Deaths	131,044	2,064
Total Population	12,335,000	524,000
Crude Mortality Rate (per 100,000)	1,062.4	393.4

152

% Distribution of Age Groups



Vita

Direct Adjustment

Age-adjusted Mortality Rates for FL and AK, 1988
Calculated by Direct Adjustment

	Age-Specific Death Rates/100,000			Expected # of Deaths	
Age (years)	Florida	Alaska	US Pop. (millions)	Florida	Alaska
<5	284	274	18.3	52,000	50,000
5 - 19	57	65	52.9	30,000	34,000
20 - 44	198	188	98.1	194,000	184,000
45 - 64	815	629	46.0	375,000	289,000
>65	4425	4350	30.4	1,345,000	1,322,000
		Total	245.7	1,996,000	1,879,000

Florida = 1,996,000/(245.7 x 10^6) = 812.4 per 100,000

Alaska= 1,879,000/(245.7 x 10^6) = 764.8 per 100,000

154

Comparing Rates

Are people more likely to die in Florida than in Alaska?

	Florida	Alaska
Number of Deaths	131,044	2,064
Total Population	12,335,000	524,000
Crude Mortality Rate (per 100,000)	1,062.4	393.4
Age Adjusted Mortality Rate (per 100,000)	812.4	764.8

155

Picking the 'Standard'

- Actual adjusted values will vary with the selection of the 'standard' population
  - Comparisons between groups will usually remain fairly constant
- Important: All groups in the comparison should be adjusted to same standard

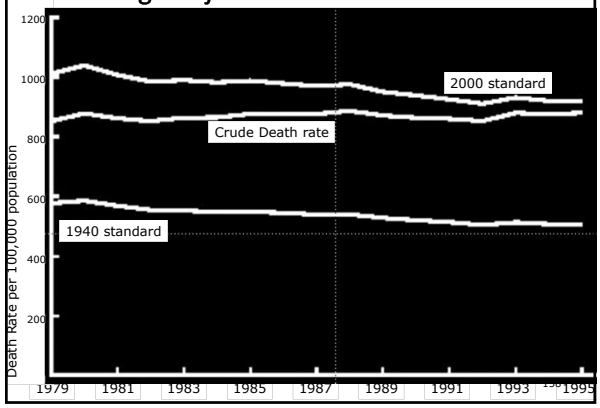
156

Standard Populations

	1940 Standard		2000 Standard	
Age	Population	Distribution	Population	Distribution
<1	2,020,174	0.015343	3,795,000	0.013818
1–4	8,521,350	0.064718	15,192,000	0.055317
5–14	22,430,557	0.170355	3,997,000	0.145565
15–24	23,921,398	0.181677	38,077,000	0.138646
25–34	21,339,026	0.162066	37,233,000	0.135573
35–44	18,333,220	0.139237	44,659,000	0.162613
45–54	15,512,071	0.117811	37,030,000	0.134834
55–64	10,572,205	0.080294	23,961,000	0.087247
65–74	6,376,189	0.048426	18,136,000	0.066037
75–84	2,278,383	0.017304	12,315,000	0.044842
85+	364,752	0.002770	4,259,000	0.015508
Total	131,669,275	1.000000	274,634,000	1.000000

157

Crude & Age-adjusted Death rates



Indirect Adjustment

- Standard population rates x observed population
  - Apply age-specific rates from standard population to
    - age-specific population under study
  - Calculate 'expected' number of cases (if the rates from a standard population were applied)
  - Then compare to what was observed:

    SMR = observed/expected x 100%

159

Indirect Standardization

		Population (millions)		Expected Deaths	
Age	US Death Rates	Florida	Alaska	Florida	Alaska
<5	251.1	.85	.06	2,134	151
5–19	47.2	2.28	.13	1,076	61
20–44	161.8	4.41	.24	7,135	388
45–64	841.9	2.60	.08	21889	674
65+	5,104.8	2.20	.02	112,306	1,021
		Total Expected		144,539	2,295
		Total Observed		131,044	2,064

SMR = o/e x 100% = 90.7% 89.9%

160

Interpreting the SMR

- < 100%
  - This population has fewer events than you would expect based on the standard rates
- = 100%
  - This population has the same number of events that you would expect based on the standard rates
- > 100%
  - This population has more events than would be expected based on the standard rates

161

When to use indirect adjustment?

- When category-specific rates are not known
- When populations are small (and rates not stable)
  - Occupational settings
  - Small communities for short time periods

162

Direct vs Indirect Adjustment

- Adjustment

  Strata pop X Strata Rate

  – Direct Adjustment
    - Standard Population X OBSERVED RATES

  – Indirect Adjustment
    - OBSERVED POPULATION X Standard Rates

163

Summary

- Different types of rates
  – Crude, category specific
- How to compare two rates?
  – Strata-specific always appropriate
  – Direct-adjustment for summary comparisons
- Direct: Standard populations
- Indirect: Standard rates

164

Interpretation: Chance

Introduction

165

Learning Objectives

- Understand the role of chance in epidemiologic studies

166

Sampling

- Goal of study
  – Determine the true relation between exposure and disease
- Actual results may vary
  – Sample vs. whole population
  – Sampling variability

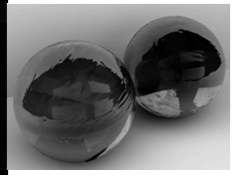
167

Statistics 101

Urn holds 100 marbles

- red and/or blue

draw 4



168

Statistics 101

Red	Blue	Conclusion
0	4	All blue
1	3	75% blue
2	2	50% blue
3	1	25% blue
4	0	All red

169

Statistics 101

50 Blue, 50 Red

Red	Blue	Probability	Conclusion
0	4	6%	All blue
1	3	25%	75% blue
2	2	38%	50% blue
3	1	25%	25% blue
4	0	6%	All red

170

Statistics 101

50 Blue, 50 Red

Sample Size	Probability of sample with 1 color (%)
4	12.0
5	5.6
6	2.7
7	1.2
8	0.6
9	0.3
10	0.1
15	0.0018
20	0.000018

171

Statistics 101

- Given
  - Hypothesis
    - the chance of drawing a red marble on any one try is 50%
  - information about sample size
- Possible to calculate probability of
  - a bad sample, or
  - observing a particular result from a set of sample data

172

Statistics 101

"Bad samples happen" – S. Pettygrove

↑ sample size →
↓ sampling variability and
↓ probability of an unrepresentative sample

173

Interpretation: Chance

Statistical Inference

174

Learning Objectives

- Understand how chance affects our interpretation of the results of an epidemiologic study

175

Epi Studies

- Measure disease frequency in two (or more) groups
  - differ only on the exposure of interest
- 2 measures of disease frequency
  - -> single measure
    - RR,
    - OR,
    - Risk or rate difference

176

Statistical Inference

- Unrepresentative sample due solely to chance
  - measure of association you observe ≠
  - true measure of association
  - by chance alone
- Possible to calculate probability that observed measure of association was due to chance

177

Hypothesis Testing

- Hypothesis testing
  - performing a statistical test
  - Particular statistical test to use depends on type of study, type of measurement, etc.
- Statistical test
  - quantifies the probability that sampling variability or chance may explain the observed association

178

Hypothesis Testing

- Null hypothesis
  - (H_0): RR=1, OR=1, RD=0
  - Assume H_0 is true
  - NOT some alternative hypothesis (H_A)
- H_0 assessed by statistical test
  - Reject H_0 or
  - Fail to reject H_0

179

T-test

Continuous Data (like height)

$$H_0: \bar{x}_1 - \bar{x}_0 = 0$$

$$t_{(df)} = \frac{\bar{x}_1 - \bar{x}_0}{\sqrt{s_p^2 \Big/ \left(1/n_1 + 1/n_2\right)}}$$

Where:
x_1 = mean of group 1
x_2 = mean of group 2
s_p = pooled estimate of variance
n_1 = number in group 1
n_2 = number in group 2

180

Chi-square

Categorical Data

$$\chi^2_{(df)} = \sum \frac{(O_i - E_i)^2}{E_i}$$

Where:
O_i = Observed in the i_{th} group
E_i = Expected in the i_{th} group
df = degrees of freedom for the statistic

181

Hypothesis Testing

- Statistical test → p-value
- P-value
  - *Probability* observed result would occur, if the null hypothesis is really the truth
  - Given that H_0 is true, the p-value is the probability of seeing the observed result, or results *more* extreme, by chance alone

182

Hypothesis Testing

- P value ranges from 0 to 1
- "Small" p-values
  - Low probability that a result as extreme (or more so) would have been observed if H_0 were true
  - H_0 incompatible with observed data
  - alternate hypothesis a better explanation for the data
  - chance is an unlikely explanation for the result

183

Statistical Conventions

- $p \leq 0.05$
  - arbitrary cutoff for statistical significance
- If $p \leq 0.05$
  - results unlikely to be due to chance
  - reject H_0 in favor of H_A

184

Statistical Conventions

- If p> 0.05
  - the finding can be explained by chance
  - do not reject H_0
- NOTE
  - Chance cannot be completely excluded as a possible explanation no matter how small the p-value

185

More on the P value….

- P values determined by
  - magnitude of the association
  - sample size (sample variability)
- Huge sample →
  - even a trivial risk increase or decrease is statistically significant
- Small sample →
  - even a large risk increase or decrease is not statistically significant

186

p-Value Example

- In utero DES Exposure and the Risk of Breast Cancer
- H_0
  - There is no association of in utero DES exposure and breast cancer risk
- H_A
  - There is an association of in utero DES exposure and breast cancer risk OR (two-sided test)
  - In utero DES exposure increases the risk of breast cancer (one-sided test)

187

p-Value Example

- RR = 1.4 p-value = .10

- Best estimate of the increased breast cancer risk associated with DES is 1.4
- p-value > 0.05
  - not "statistically significant"

188

p-Values

- p-value
  - guide to whether chance is an explanation
- Statistical significance
  - evaluates only the role of chance
  - Does not test for bias or confounding
- Statistical significance vs biological significance
- Multiple testing effects

189

Interpretation: Chance

Confidence Intervals

190

Learning Objectives

- Understand the concept and interpretation of confidence intervals

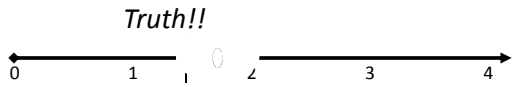
191

Confidence Intervals

- Means of quantifying sampling variability
- Measure of association (RR, OR) is the point estimate
- Point estimate has variability
  - can be expressed mathematically

192

Chance



Measure of Association

193

Confidence Intervals

- Formulas to calculate CIs specific to measure of association
  - See biostat books / class
- What goes into the calculation?
  - Size of point estimate
    - Point estimate ± a measure of variance
  - Sample size
    - ↑ sample size → ↓ size of CI
  - Z score
    - 95% CI : α = 0.05 : 1.96

194

Confidence Intervals

- RR and 95% CI
  - 1.6 (1.5 - 1.7)
  - 1.7 (1.3 ---------- 2.5)
  - 2.5 (0.95 ---------- 6.5)

195

Confidence Intervals

Increase sample size

Decrease variance

Increase α
(from 95% CI to 60% CI)

196

Things to Remember

- CI is a measure of variability around the point estimate
  - The point estimate will ALWAYS fall within the CI
- CI does indicate a range of values that are consistent with the data
  - If the CI includes 1, you cannot reject the null hypothesis of no association

197

Interpreting CIs

"The 95% confidence interval means that if you repeated the study 100 times, and calculated a confidence interval each time, you expect that the true value of the measure of association (RR or OR) would fall within that interval for 95 of those repetitions."

» Sylvan Green

198

Interpreting CIs

- Note
  - In 5 out of the hundred repetitions the TRUE value would be outside the 95% CI
- Note also
  - You most likely did the experiment ONCE
  - Is it one of the 95 or one of the 5?

199

p-values vs CIs

- p-value
  - null hypothesis compatible with the data?
    - Yes / No
- CI
  - range of hypotheses that are compatible with the data
    - Size of CI → how much weight to place on results
    - "CI you could drive a truck through"

200

CIs Example

- DES and breast cancer
  - RR = 1.4
    - 1.4 = best estimate of the increased breast cancer risk associated with DES
  - 95% CI = 0.7 - 2.6
  - Data consistent with an association of sizes 0.7 to 2.6
  - Data consistent with RR = 1
  - Cannot reject the hypothesis of no increase in breast cancer risk with DES exposure

201

Statistical Testing

- p-Values and CIs tell you nothing about other possible explanations for an observed result
  - Bias?
  - Confounding?
- p-Values and CIs tell you nothing about biological, clinical or public health significance

202

Horse Race Analogy

- Betting on a horse
  - which horse
  - how much money
- Point estimate is which horse to bet on
  - DES and breast cancer example
    - best bet for a winning RR is 1.4

203

Horse Race Analogy

- P value and CI
  - how much money to bet
    - p-value relatively large
    - CI wide
  - Do not bet a lot of money on this "horse"
- If small P value and narrow CI, *bet more*

204

Interpretation: Chance

Design Issues & Summary

205

Learning Objectives

- Understand issues of chance in epidemiology and how to account for them in the design phase

206

Chance in Study Design

- Design effect
  - sample size
  - power

207

Estimating Sample Size

- You will need
  - Probable size of the effect
  - Proportion of exposure in the control group
  - Size of the Type I error you are concerned about (alpha level)
  - Size of the Type II error you are concerned about (power)

208

Errors in Inference

Significance Testing

	Truth: H_0 True	Truth: H_A True
Do not Reject H_0 (not significant)	Correct	Type II
Reject H_0 (significant)	Type I	Correct

209

Chance in Study Design

- Type I error = p value
  - α
- Type II error = β (beta) error
  - 1- β = Power

210

Example: Study 1

Blood Test	Males	Females
Positive	52	48
Negative	48	52

OR = 1.17 95% CI = (0.65 – 2.12)

211

Example: Study 2

Blood Test	Males	Females
Positive	5200	4800
Negative	4800	5200

OR = 1.17 95% CI = (1.11 – 1.24)

212

Example

Study A

- OR = 2.6 95% CI = 1.9 – 4.2

Study B

- OR = 2.6 95% CI = 0.54 – 17.2

Either

- Sample size too small – insufficient power

or

- No true relationship

213

Exercise

- Interpreting p values and confidence intervals
- Five studies
  - same exposure-disease relationship
  - assume no bias or confounding

214

Exercise

Study	n	RR	p-value	95% CI
A	100	3.1	.10	0.8 - 4.2
B	500	3.0	.06	0.9 – 3.3
C	1000	3.5	.02	2.6 – 4.5
D	2000	3.2	.015	2.2 – 3.5
E	2500	3.3	.001	2.8 – 3.6

215

Interpretation

Bias vs.. Confounding & Other Issues

216

Learning Objectives

- Some wrap-up notes on bias and confounding

217

Bias vs.. Confounding

- Difference between confounding and bias:
  - you collected data on the confounder
    - you can examine it and
    - control for it in analysis
  - No data collected on biasing variable
    - nothing you can do about it once data collected

- paraphrased from Dr. Moyses Sklo, Johns Hopkins University

218

Confounder Criteria

1. It must *NOT* be in the causal pathway
2. It MUST be associated with the exposure of interest
3. It MUST be associated with the outcome
4. It must *NOT* be an effect modifier
5. It MUST affect the measure of association between the exposure of interest and the outcome (i.e., $OR_{crude} \neq OR_{strata}$)

219

Bias

- If it affects both groups equally then it is NOT BIAS
  - Study subjects are all 103 years of age or older
    - Recall is definitely likely to be a *problem*
  - It is only recall BIAS if the problem affects the two groups (exposed/unexposed or diseased/healthy) differently

220

Writing about Bias

- Hypothetical test question:

What are potential sources of bias for this study?

A **bad** answer: "Recall bias"

A **good** answer: "Recall bias may be a problem because cases were interviewed within 2 days but the average time to interview was 14 days for controls"

A **better** answer: The good answer PLUS "This would be likely to artificially inflate the measure of association <u>because</u> cases will be more likely than controls to remember their exposure"

221

Validity vs. Generalizability

- The study is
- Valid
  - If the association between exposure and disease has been accurately assessed
- Generalizable to the population represented by the study subjects

222

Validity vs. Generalizability

- A study may be Valid within a very specific population
  - Left-handed, Norwegian, vegetarian, women
- But not Generalizable to the general population

223

Validity vs. Generalizability

- If the study is not Valid then who cares whether it is Generalizable or not?

224

Target Population

- The population to which you would like to Generalize the results:
  - Ambulatory, community-living, adults ages 65+ (footwear study)
  - People with age-associated memory-impairment (gingko study)

225

Screening

226

Learning Objectives

- Understand the public health context and considerations in screening

227

Screening

- Definition
  - Presumptive identification of unrecognized disease or defect by the application of tests, examinations, or other procedures which can be applied rapidly
- How does a screening test differ from a diagnostic test?

228

Screening

- 2^0 prevention
  - goal is to reduce morbidity and mortality
- Reduce burden on the individual
  - potential years of life lost
  - extent of disability, pain and discomfort
  - cost of treatment
  - impact on individual and family
- Reduce burden on society
  - Mortality
  - Morbidity
  - Societal costs of treatment

229

Screening Considerations

- When do you consider screening?
  - Characteristics of Disease
    - Disorder is well-defined
    - Pre-clinical phase
    - Prevalence is known
    - Effective treatment readily available*
    - Follow-up & treatment for a positive result are agreed upon, acceptable to patients, and available

230

Screening Considerations

- When do you consider screening?
  - Characteristics of Test
    - Parameters of test
      - Sensitivity & specificity, reliability
    - Cost
    - Acceptability / safety of test
    - Cost-effective
    - Simple and safe
    - Facilities are available or easily installed
    - Simple to perform, easy to interpret, capable of use by various health professionals

231

Screening Considerations

- When do you consider screening?
  - Public Health Context
    - Equity – available to all
    - Consequences of False Positives
      - Intrusiveness /cost of follow-up
    - Consequences of False Negatives
    - Cost / Benefit analysis
    - Benefit vs Harm from test
    - Diversion of health resources to new screening tests and follow-up
    - Society values for new screening test vs other health services
- PSA

232

Evaluating Screening

- Reduction of mortality
- Reduction of case fatality in screened vs unscreened
- Reduction of complications
- Prevention / reduction of recurrences
- Improvement in quality of life indicators

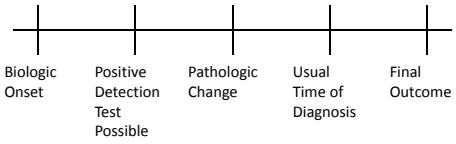
233

Screening Problems

- Referral bias: selection in types who are screened – also called volunteer bias
- Length-biased sampling: selection bias in the type of disease detected among those who are screened
- Lead time bias

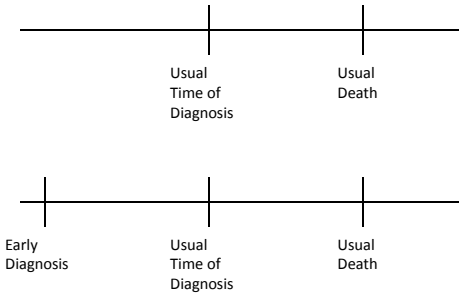
234

Screening



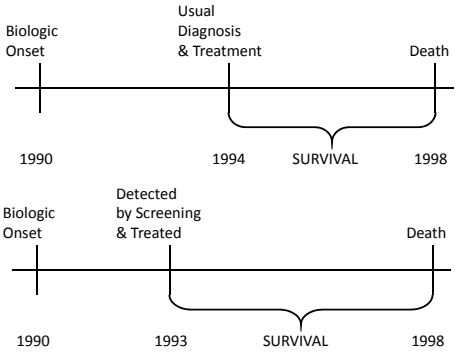
235

Screening



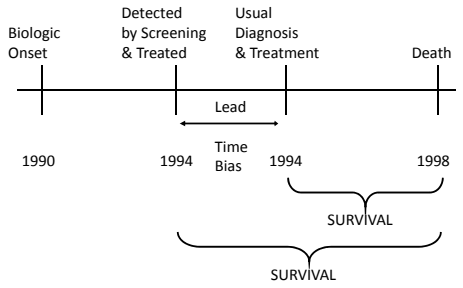
236

Screening



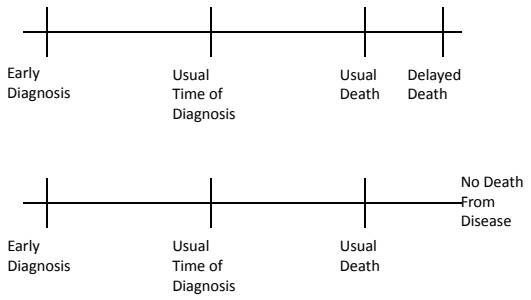
237

Screening



238

Screening



239

Surveillance

240

Learning Objectives

- Understand the purpose and types of surveillance systems

241

Surveillance

- Definitions
  - "…continued watchfulness over the distribution and trends of incidence through the systematic collection, consolidation, and evaluation of morbidity and mortality reports and other relevant data with timely and regular dissemination to those who 'need to know'…"
    - » Langmuir, 1963, NEJM, 268, 182-192
  - Systematic ongoing collection, collation, and analysis of data and the timely dissemination of information to those who need to know so that action can be taken.
    - » World Health Organization
  - The ongoing systematic collection, analysis, and interpretation of health data, essential to the planning, implementation, and evaluation of public health practice, closely integrated with the timely dissemination of these data to those who need to know. The final link in the surveillance chain is the application of these data to prevention and control. A surveillance system includes a functional capacity for data collection, analysis and dissemination linked to public health programs.
    - » Centers for Disease Control and Prevention

242

Surveillance

- Key Components
  - Data collection
  - Analysis
  - Dissemination
    - *Action!*

243

History of Surveillance

- John Gaunt—1662—first to record mortality data
- Johann Peter Frank—1776—first public health monitoring of schoolchildren's health, prevention of injuries, maternal and child health, and public water and sewage disposal
- William Farr—1839—Father of modern surveillance—developed vital statistics and weekly, quarterly, and annual reporting
- United Kingdom—1899—first compulsory notification of selected infectious disease
- United States—1878—national morbidity data on plague, smallpox, and yellow fever
- United States—1925—all states reported weekly to the USPHS on selected diseases
- Worldwide—1907-1950s—from Europe to most of the globe, mortality reporting was required, as was surveillance of selected communicable diseases

244

Surveillance Objectives

1. To characterize disease patterns by time, place, and person
   - To suggest etiologic hypotheses
2. To detect epidemics (e.g., communicable diseases)
   - To identify bioterrorist events
3. To detect changes in incidence (e.g., chronic diseases)
4. To identify cases for analytic epidemiologic research
5. To evaluate prevention and control programs
6. To project future health care needs

245

Surveillance Types

- Passive
  - Required reporting by health professionals
  - Passive on the part of those running the system, but requires action from health professionals
- Active
  - Those administering the system send teams out to data sources to actively look for cases

246

Passive Surveillance - Reportable Diseases in AZ

Amebiasis
Anthrax
Aseptic meningitis: viral
Basidiobolomycosis
Botulism
Brucellosis
Campylobacteriosis
Chancroid (*H. ducreyi*)
Chlamydia infection, genital
Cholera
Coccidioidomycosis (Valley Fever)
Colorado tick fever
Conjunctivitis: acute (outbreaks)
Creutzfeldt-Jakob disease
Cryptosporidiosis
Cyclospora infection
Cysticercosis
Dengue
Diarrhea, nausea, or vomiting (outbreaks)
Diphtheria
Ehrlichiosis
Emerging or exotic disease
Encephalitis: Viral or parasitic
Enterohemorrhagic *Escherichia coli*
Enterotoxigenic *Escherichia coli*
Giardiasis
Gonorrhea
Haemophilus influenzae: (invasive)
Hansen's disease (Leprosy)
Hantavirus infection
Hemolytic uremic syndrome
Hepatitis A
Hepatitis B and D
Hepatitis C
Hepatitis E
Herpes genitalis
Human Immunodeficiency Virus (HIV) & related disease
Kawasaki syndrome
Legionellosis (Legionnaires' Disease)
Leptospirosis
Listeriosis
Lyme disease
Lymphocytic choriomeningitis
Malaria
Measles (rubeola)
Meningococcal Invasive Disease
Mumps Pertussis (whooping cough)
Plague
Poliomyelitis
Psittacosis (ornithosis)
Q Fever
Rabies in a human
Relapsing Fever (borreliosis)
Reye Syndrome
Rocky Mountain Spotted Fever
Rubella (German measles)
Rubella syndrome, congenital
Salmonellosis
Scabies (outbreaks only)
Severe acute respiratory syndrome (SARS)
Shigellosis
Smallpox
Streptococcal Group A: invasive Streptococcal Group B: invasive in infants <90 days of age
Streptococcus pneumoniae (invasive)
Syphilis
Taeniasis
Tetanus
Toxic Shock Syndrome
Trichinosis
Tuberculosis
Tuberculosis infection (child < 6yrs) Tularemia
Typhoid fever
Typhus fever
Unexplained death + history of fever
Vaccinia-related adverse event
Vancomycin-resistant Enterococcus sp.
Staphylococcus aureus (Vancomycin-resistant / intermediately susceptible)
Varicella (chickenpox)
Vibrio infection
Viral hemorrhagic fever
West Nile virus infection
Yellow fever
Yersiniosis

http://azdhs.gov/phs/oids/rptlist.htm#When%20should%201%20report

247

Arizona Administrative Code* Requires Providers To:

REPORT 'EM

to the Local Health Department Phone#

☐2*O	Amebiasis	☐	Hantavirus infection	☐2*O	Salmonellosis
☒	Anthrax	☒	Hemolytic uremic syndrome	O	Scabies
☐	Aseptic meningitis: viral	☐2*O	Hepatitis A	☒	Severe acute respiratory syndrome
☐	Basidiobolomycosis	☐	Hepatitis B and D	☐2*O	Shigellosis
☒	Botulism	☐	Hepatitis C	☒	Smallpox
②	Brucellosis	☐2*O	Hepatitis E	☐	Streptococcal Group A: Invasive disease
☐2*O	Campylobacteriosis	☐	Herpes genitalis	☐	Streptococcal Group B: Invasive disease in infants younger than 90 days of age
☐	Chancroid	☐	HIV infection and related disease	☐	*Streptococcus pneumoniae* (pneumococcal invasive disease)
☐	*Chlamydia* infection, genital	☐	Kawasaki syndrome	☐	Syphilis
②*	Cholera	☐	Legionellosis (Legionnaires' disease)	☐2*O	Taeniasis
☐	Coccidioidomycosis (valley fever)	☐	Leptospirosis	☐	Tetanus
☐	Colorado tick fever	☒	Listeriosis	☐	Toxic shock syndrome
O	Conjunctivitis: acute	☐	Lyme disease	☐	Trichinosis
☐	Creutzfeldt-Jakob disease	☐	Lymphocytic choriomeningitis	②	Tuberculosis
☐2*O	Cryptosporidiosis	☐	Malaria	②	Tuberculosis infection in a child younger than 6 (positive test result)
☐	*Cyclospora* infection	☒	Measles (rubeola)	☒	Tularemia
☐	Cysticercosis	☒	Meningococcal invasive disease	☒	Typhoid fever
☐	Dengue	②	Mumps	②	Typhus fever
O	Diarrhea, nausea, or vomiting	☒	Pertussis (whooping cough)	☒	Unexplained death with a history of fever
☒	Diphtheria	☒	Plague	②	Vaccinia-related adverse event
☐	Ehrlichiosis	☒	Poliomyelitis	☐	Vancomycin-resistant *Enterococcus* spp.
☒	Emerging or exotic disease	☐	Psittacosis (ornithosis)	☒	Vancomycin-resistant or Vancomycin-intermediately susceptible *Staphylococcus aureus*
②	Encephalitis, viral or parasitic	②	Q fever	☒	Vancomycin-resistant *Staphylococcus epidermidis*
☒	Enterohemorrhagic *Escherichia coli*	☒	Rabies in a human	☐	Varicella (chickenpox)
☒	Enterotoxigenic *Escherichia coli*	☐	Relapsing fever (borreliosis)	☐2*O	*Vibrio* infection
☐2*O	Giardiasis	☐	Reye syndrome	☒	Viral hemorrhagic fever
☐	Gonorrhea	☐	Rocky Mountain spotted fever	☒	West Nile virus infection
☐	*Haemophilus influenzae*: invasive disease	②*	Rubella (German measles)	☒	Yellow fever
☐	Hansen's disease (Leprosy)	②	Rubella syndrome, congenital	☐2*O	Yersiniosis

☒ Submit a report by telephone or through an electronic reporting system authorized by the Department within 24 hours after a case or suspect case is diagnosed, treated, or detected or an occurrence is detected.

* If a case or suspect case is a food handler or works in a child care establishment or a health care institution, instead of reporting within the general reporting deadline, submit a report within 24 hours after the case or suspect case is diagnosed, treated, or detected.

② Submit a report within one workday day after a case or suspect case is diagnosed, treated, or detected.

☐ Submit a report within five workday days after a case or suspect case is diagnosed, treated, or detected.

O Submit a report within 24 hours after detecting an outbreak.

A.A.C. R9-6-202, 301

Effective 10/07/2004

Surveillance

- From local to global
  - Pima County
  - Arizona
  - United States
  - Selected Populations
  - Pan America
  - World

249

Surveillance

- Arizona Department of Health Services (ADHS) (www.hs.state.az.us)
  - Communicable diseases
  - Cancer registry
  - Bioterrorism
  - West Nile Virus

250

ADHS Cancer Registry

- The Arizona Cancer Registry
  - population-based surveillance system
    - incidence and
    - survival of persons having been diagnosed with cancer
- Arizona Revised Statute §36-133
  - Mandates the reporting of cancer cases in the state of Arizona
  - Required to report
    - hospitals,
    - clinics, and
    - physicians

251

ADDSP

- Arizona Developmental Disabilities Surveillance Program
- UA Pediatrics
- CDC funding
- System-wide protocol

252

What is Autism?

- Autism 1943 Kanner
- Kanner L. Autistic disturbances of affective contact. Nervous Child 2:217-250.
  - Deficits in 3 areas
    - Social
    - Communication
    - Behavior and interests
  - Approximately 70% have IQ's < 70 (mental retardation) – *as originally described…*

253

Diagnostic Criteria

- "There are no definitive diagnostic tests for autism. Diagnosis is made from a detailed developmental history and observation of behavior in structured and unstructured situations. This process is fraught with difficulties of definition and standardiz(s)ation."

- *Wing and Potter, 2002. Ment Ret and Dev Dis Res Rev*

254

Autism

- Autistic Disorder
  - At least 6 criteria from 3 areas:
    - Impaired social interaction (at least 2)
    - Impaired communication (at least 1)
    - Activities, behavior and interests that are repetitive, restricted and stereotyped (at least 1)
  - Symptoms not better explained by Childhood Disintegrative Disorder or Rett's Disorder

Differences are QUALITATIVE, not only the result of delays

Development in these areas follows a DIFFERENT path than that of most children.

255

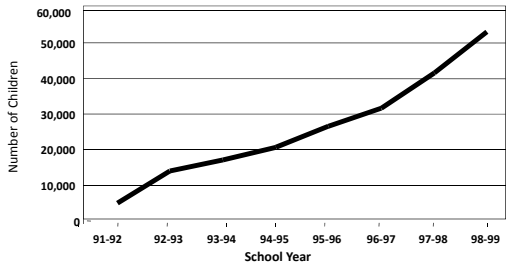
Individuals with Disabilities Education Act

- 1991 IDEA Part B
  - Mandates that the public schools provide a free and appropriate education for all children with disabilities between the ages of 3 and 21

256

CA Service Data

Children (ages 6-17 years) Served for Autism Under IDEA



Source: U.S. Department of Education, Office of Special Education Programs, Data Analysis System (DANS)

257

CA Service Data

- 1987-1998 Data
  - Autistic Disorder: >200% ↑
  - ASD: 1966% ↑
  - Other Developmental Disabilities: 31-49% ↑

258

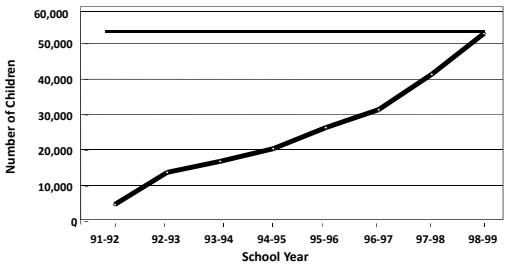
California Service Data

- Increase from 3,795 to over 53,000 children served under autism classification
  - does not consider population growth
- Switching from services for mental retardation to autism
  - Croen LA, Grether JK, Hoogstrate J, and Selvin S. The changing prevalence of autism in California. J Autism Dev Disord. 2002;32:207-215.

259

CA Service Data

Children (ages 6-17 years) Served for Autism



Source: U.S. Department of Education, Office of Special Education Programs, Data Analysis System (DANS)

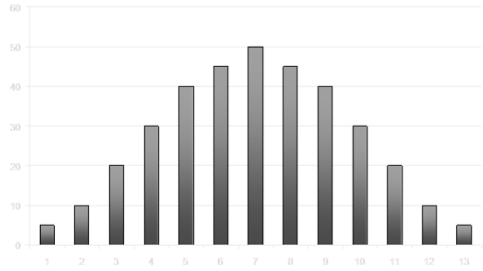
260

Fuzzy Outcomes

- Autism
- High blood pressure
- High blood sugar
- Obesity

261

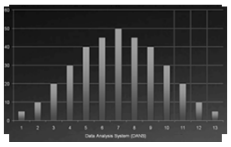
Normally Distributed Traits



262

Artifactual Prevalence Increase

- Increased availability of services
- Expanding case definition
- ↑ Awareness among parents and providers
- ↓ stigma
- ↓ age @ diagnosis
- Changing ascertainment methodology
- Diagnostic patterns



Determining Disease Frequency

- Define the population
- Define a case *
- Identify sources of information
- Count cases using a systematic approach *

264

ADDM NET

Autism and Developmental Disabilities Monitoring Network

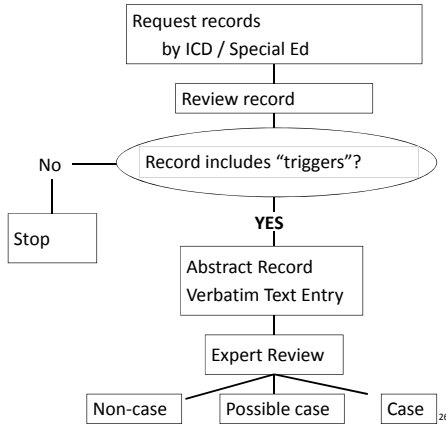
265

Data Sources

- Special Educational Records
  - All public schools in Maricopa Co.
- Clinical Records
  - Phoenix Children's Hospital
  - St. Joseph's Hospital Developmental Clinic

266

Study Procedures



267

Data Collected

- Name
- Date of birth
- State of birth
- Gender
- Race/ethnicity
- Parent names, address/phone to determine residency for surveillance year
- Special ed exceptionalities
- Evaluation Information – date, examiner, reason for referral, verbatim behaviors and diagnoses
- MR/IQ, Adaptive, and ASD Test information

268

Autism Results

- Morbidity and Mortality Weekly Report
  - (www.cdc.gov/mmwr)
  - Weekly bulletin on surveillance
  - Electronic Subscription available free
- Example: lead poisoning in children
  - Report includes: definition of the condition
  - Reporting period
  - State and local surveillance
  - Description of CDC prevention programs
- National surveys
- Extensive bibliography
- Maps and graphs
- Results for reporting period

269

Registry vs. Surveillance

- How to distinguish between a registry and a surveillance system
  - Registry
    - Consent
    - Contact
  - Surveillance
    - Waiver of consent
    - No contact

270

Surveillance Concerns

- Emerging Issues
  - SARS
  - West Nile Virus
  - Influenza
  - Tuberculosis

271

Agencies Conducting Surveillance

- Pan American Health Organization (PAHO)
  - Basic Country health profiles for the Americas
  - Health statistics from the Americas
  - Identification of and connection to each country's epidemiology and surveillance system
  - EID weekly updates: emerging and reemerging infectious diseases and other conditions
    - Antimicrobial resistance
    - Campylobacter
    - Chaga
    - Cholera
    - Dengue
    - Malaria
    - Salmonella
    - Shigella
    - Tuberculosis
    - SARS
    - Veterinary Public Health
    - Non-Communicable Diseases

272

Agencies Conducting Surveillance

- National Institute for Occupational Health and Safety (NIOSH)
  - Data collection to fill gaps in other surveillance data
    - Occupational health surveillance tracks
      - injury
      - illness
      - hazards
      - exposures
  - Began with Occupational Safety and Health Act (OSHA) of 1970

273

Surveillance Links

- Agencies
  - Centers for Disease Control and Prevention (CDC)
    - www.cdc.gov
  - National Institute for Occupational Safety & Health
    - www.cdc.gov/niosh/topics/surveillance
  - National Center for Health Statistics
    - www.cdc.gov/nchs
  - Pan American Health Organization
    - www.paho.org

274

Surveillance Links

- Data and Statistics
  - National Electronic Disease Surveillance System (NEDSS) - http://www.cdc.gov/nedss
    - http://www.cdc.gov/scientific.htm with links to:
      - CDC and ATSDR Electronic Information Resources for Health Officers
      - CDC WONDER
      - Hazardous Substance Release / Health Effects Database (HAZDAT)
      - HealthComm Key
      - Injury Maps
      - WISQARS – Web-based Injury Statistics Query and Reporting System
      - 121 Cities Mortality Reporting System
      - Assisted Reproductive Technology Success Rate
      - Behavioral Risk Factor Surveillance System
      - Birth Defect Surveillance

275

Surveillance Links

- More Data and Statistics….
  - http://www.cdc.gov/scientific.htm with links to:
    - Cancer Registries Program
    - Hazardous Substances Emergency Events Surveillance
    - HIV/AIDS Surveillance Report
    - National Notifiable Diseases Surveillance System (NNDSS)
    - National Oral Health Surveillance System
    - Pediatric Nutrition Surveillance System
    - Pregnancy Risk Assessment Monitoring System
    - Sexually Transmitted Diseases
    - Surveillance Resources for Infectious Diseases
    - Tuberculosis Surveillance Reports
    - Workplace Safety and Health Surveillance
    - Youth Risk Behavior Surveillance System
    - Laboratory Practice Standards Information

276