Quantitative Research Methods Training

Sponsored by the Chickasaw Nation and the Oklahoma Shared Clinical and Translational Resources

Ashley White, MPH and Julie Stoner, PhD
University of Oklahoma Health Sciences Center

1

Acknowledgements

- Julie Stoner, PhD
- Sydney Martinez, PhD
- Amanda Janitz, PhD
- Michael Peercy, MPH, MT(ASCP)H
- Bobby Saunkeah, RN, MSHCE, CIP
- Tasha Matthews

- Some of these materials were developed as part the Association of American Indian Physicians' Data Into Action training, funded by the Centers for Disease Control and Prevention through the support of Cooperative Agreement Number 1U38OT000133.
- CDC Principles of epidemiologydemiology in Public Health Practice https://www.cdc.gov/ophss/csels/dsepd/.html

Grant support provided through National Institutes of Health, National Institute of General Medical Sciences [Grant 2U54GM104938-06, PI Judith James] and National Institutes of Health, National Institute on Minority Health and Health Disparities [Grant 1R25MD011564, PI Julie Stoner/Courtney Houchen].

2

Objectives

1. Define epidemiology and epidemiologic measures of disease burden
2. Define and interpret descriptive statistics and charts
3. Identify pitfalls in data interpretation and data summaries
4. Use Excel to create data tables and charts

SCTR Oklahoma Shared Clinical & Translational Resources

3

ACTIVITY 1:
What are you doing with data?
Does it affect your everyday life?

SCTR Oklahoma Shared Clinical & Translational Resources

4

Epidemiology in the news

The weed-killing chemical involved in a Monsanto lawsuit was found in Cheerios and Quaker Oats products. Here's how worried you should be.

SCTR Oklahoma Shared Clinical & Translational Resources

5

EPIDEMIOLOGY HISTORY, DEFINITION, AND USES

SCTR Oklahoma Shared Clinical & Translational Resources

6

John Snow: Pioneer of Epidemiology

https://oeta.pbslearningmedia.org/asset/envh10_vid_johnsnow/

SCTR Oklahoma Shared Clinical & Translational Resources

7

Epidemiology defined

- The study of factors affecting the health and illness of populations, and serves as the foundation and logic of interventions made in the interest of public health and preventive medicine
  - Wikipedia
- The study of the distribution and determinants of health related states in specific populations, and the application of this study to control of health problems
  - A Dictionary of epidemiology

SCTR Oklahoma Shared Clinical & Translational Resources

8

Roles of epidemiology

- **Monitor**
  - Population health status, its determinants, and its threats
- **Diagnose and Investigate**
  - Health problems and hazards
  - Determinants of health and disease
  - Disparities and inequities
  - Methods to prevent disease and promote health
- **Evaluate programs**

SCTR Oklahoma Shared Clinical & Translational Resources

9

Role of epidemiology

- **Count** cases or health events, and describe in terms of time, place, and person.
- **Divide** the number of cases by an appropriate denominator to calculate rates.
- **Compare** these rates over time or for different groups of people.

SCTR Oklahoma Shared Clinical & Translational Resources

10

Role of epidemiology

Can help to answer questions like these:

- **Who** is getting sick?
- **Where** are they getting sick?
- **When** are they getting sick?
- **What** is making them sick?
- **How** are they getting sick?
- **Why** are they getting sick?

SCTR Oklahoma Shared Clinical & Translational Resources

11

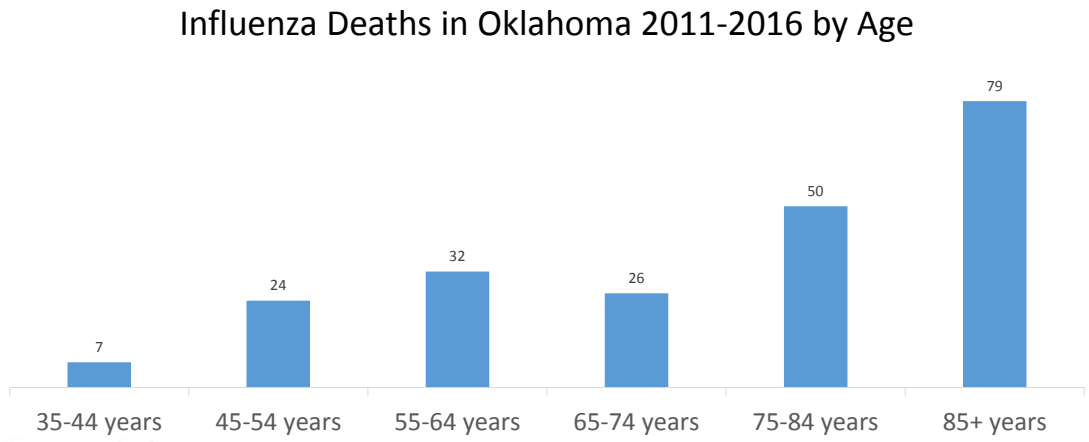
Who?

- Gender
- Age
- Race/ethnicity
- Religion
- Socioeconomic status (SES)
- Behaviors
- Occupation
- …

SCTR Oklahoma Shared Clinical & Translational Resources

12

Disease displayed by age

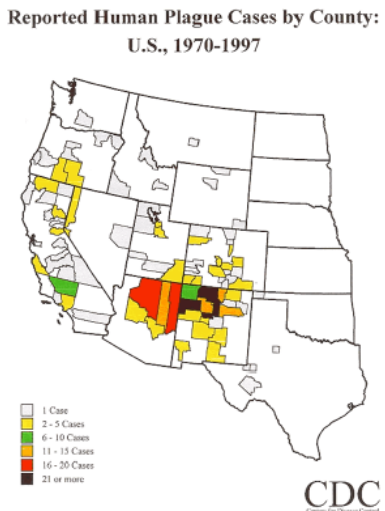


SCTR Oklahoma Shared Clinical & Translational Resources

13

Where?

- State?
- County?
- Tribe?
- Hospital?
- …



SCTR Oklahoma Shared Clinical & Translational Resources

14

When?

- Days
- Months
- Years
- Seasons
- Events
- How long an individual is sick
- How long an individual is infectious
- …

SCTR Oklahoma Shared Clinical & Translational Resources

15

Disease displayed by time and place

https://www.cdc.gov/flu/weekly/usmap.htm

SCTR Oklahoma Shared Clinical & Translational Resources

16

TYPES OF EPIDEMIOLOGIC STUDIES

SCTR Oklahoma Shared Clinical & Translational Resources

Types of epidemiologic studies

- **Non-experimental**
  - Ecologic study
  - Cross-sectional studies
  - Case-control studies
  - Cohort studies (prospective or retrospective)
  - …and many others!!!
- **Experimental**
  - Clinical trials
  - Field trials
- **Community intervention**

SCTR Oklahoma Shared Clinical & Translational Resources

Ecologic studies

- **Group/aggregate level data on exposure and outcome**
- **Example:**
  - Percent of households with guns and homicide rate among 13 countries
- **Advantages**
  - Quick and inexpensive
  - Useful for hypothesis generating
- **Limitations**
  - Cannot infer to the individual
  - Information on confounders not usually available
  - Doesn't demonstrate causal relationship

SCTR Oklahoma Shared Clinical & Translational Resources

Koepsell & Weiss, 2003

19

Cross-sectional studies

- **Prevalence Study** – Study using a cross section or sample of a population.
- **Examples**
  - Community health assessments; Behavioral Risk Factor Surveillance System
- **Advantages**
  - Measured at individual level
  - Often sample general population (generalizable)
  - Short time needed for study
- **Limitations**
  - Not good for rare diseases or diseases of short duration
  - Prevalence data (not incidence data)
  - Difficult to determine temporality/causal relationship

SCTR Oklahoma Shared Clinical & Translational Resources

Koepsell & Weiss, 2003; CDC

20

Case-control studies

- Select cases and controls and compare past exposures
- **Example**
  - National Birth Defects Prevention Study
- **Calculate Odds Ratio**
- **Advantages**
  - Rare diseases, common exposures
  - Quicker and cheaper than cohort study
- **Limitations**
  - More potential sources of bias than cohort (different types of bias)
  - Difficult to establish temporality if use prevalent cases
  - Appropriate control group difficult to find

SCTR Oklahoma Shared Clinical & Translational Resources

Koepsell & Weiss, 2003; NBDPS

21

Cohort studies

- Follow exposed and unexposed individuals over time for outcome
  - Prospective vs. retrospective
  - Calculate odds ratio, risk ratio, rate ratio
- **Example:** Framingham Heart Study
- **Advantages**
  - Better establish temporality
  - Assess multiple outcomes
  - Incidence data
- **Limitations**
  - Expensive and time consuming

SCTR Oklahoma Shared Clinical & Translational Resources

Koepsell & Weiss, 2003; Framingham Heart Study

22

Clinical trials

- Experimental Study: Study factor assigned by investigator
  - Study intervention vs. placebo or standard of care
  - Randomization or conducted in clinical setting
- Example: Risk of opportunistic infection among patients with recent HIV infection treated with either zidovudine or a placebo
- Advantages
  - Control of confounding and bias
  - Prospective
  - Most rigorous method in human populations
- Limitations
  - Not feasible for all study questions
  - Bias due to participant non-compliance
  - High cost
  - Setting is artificial

Rothman, 2002

OSCTR Oklahoma Shared Clinical & Translational Resources

23

How would you design a case-control study of the association between cardiovascular disease and cigarette smoking?

How would you design a cohort study of the association between cardiovascular disease and cigarette smoking?

How would you design a study to determine the effectiveness of a new smoking cessation program on cardiovascular disease prevention?

24

DEVELOPING MEASURES

SCTR Oklahoma Shared Clinical & Translational Resources

25

Developing measures

1. **Define the event of interest**
   Ex: diagnosis of TB, eating recommended number of vegetables, using seat belt.
2. **Define population of interest**
   Ex. Oklahomans, teens 15-19, individuals who have attempted suicide
3. **Count number of events and size of population**
4. **Time period of interest**
   Ex. years, months, weeks
5. **Constant (K)** use to put in perspective and makes various measures easier to understand

SCTR Oklahoma Shared Clinical & Translational Resources

26

Measures of frequency

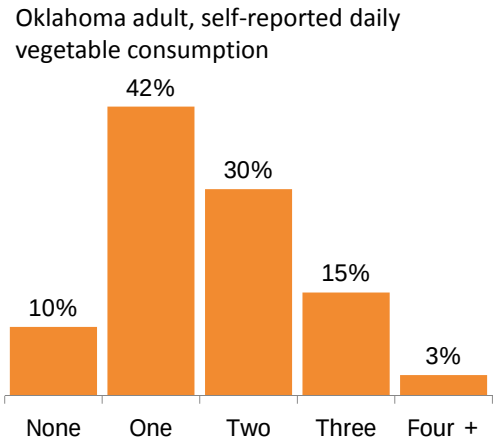
- **Proportion:** ratio in which the numerator is included in the denominator
  - **percentage** is proportion*100 (ie .25 proportion =25%)
- **Ratio measure:** one subgroup in a population divided by another subgroup
- **Rate measure:** frequency of defined events in specified population for given time period

Measures of frequency

- **Constant, K:** arbitrary number included in rate for expressing and comparing rates in similar population units
- **Average population:** Population size at mid-point of the time period.
- **Person-time:** Combines population at risk with time.

Percentages

- ¼=0.25=25%
- ½ =0.50=50%



SCTR Oklahoma Shared Clinical & Translational Resources

29

A note about converting fractions or rates

- 0.159
  - 15.9 percent (per 100)
  - 1590 per 10,000

SCTR Oklahoma Shared Clinical & Translational Resources

30

Measuring burden of disease

- Morbidity
  - Incidence
  - Prevalence
- Mortality
  - Total Mortality Rate
  - Cause Specific Mortality Rate
  - Case Fatality Rate

SCTR Oklahoma Shared Clinical & Translational Resources

31

Morbidity (illness) measures

Incidence

- Number of events per number of people per period of time (usually one year)
- May be reported per **person year as *incidence density***
- Usually refers to **new** cases of illness or disease
- Can be percent or proportion

Prevalence

- Number of events per number of people per period of time (usually one year)
- Usually refers to **existing** cases of illness or disease
- Can be percent or proportion

SCTR Oklahoma Shared Clinical & Translational Resources

32

Incidence proportion

- Number of new cases of a disease or condition in a population over a specific time
  - Ex: # new cancer cases, # youth begin smoking

$$I = \frac{Number\ of\ NEW\ cases}{Total\ population\ at\ risk}$$

SCTR Oklahoma Shared Clinical & Translational Resources

33

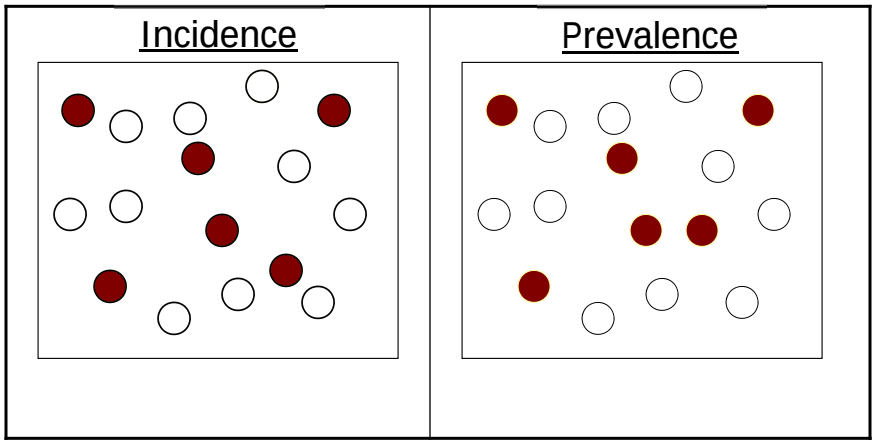
Prevalence

- Number of people with a disease or condition at a specific time in a population
  - Ex: # with diabetes, # of smokers, # of colds, # cancer patients

$$P = \frac{Number\ of\ new\ and\ existing\ cases}{Total\ population}$$

SCTR Oklahoma Shared Clinical & Translational Resources

34



SCTR Oklahoma Shared Clinical & Translational Resources

35

Calculating incidence and prevalence

- The fictitious town of Neverland has 50,000 residents. In 2012, 782 people were diabetic, and 56 people were diagnosed with diabetes
- What is the incidence and prevalence of diabetes for people who live in Neverland in 2012?

SCTR Oklahoma Shared Clinical & Translational Resources

36

Calculating incidence and prevalence

- **Incidence**:
  - 56 new cases/50,000 people/year
  - 0.00112 cases/person/year or 112 cases per 100,000 people per year
- **Prevalence**
  - (782+56)/50,000
  - 0.01676 or 1.68% or 1,676 per 100,000

SCTR Oklahoma Shared Clinical & Translational Resources

37

Mortality (death) rates

Mortality rate = $\frac{\text{(\# deaths due to certain cause)}}{\text{(\# of people in } \textbf{population of interest}\text{)}}$

- All causes
- Grouped causes (cancers)
- Specific causes (lung cancer)

SCTR Oklahoma Shared Clinical & Translational Resources

38

Mortality data

- Source – Death Certificates
  - Immediate cause
  - Contributory cause
  - Underlying cause
- Comparisons of Mortality Rate
  - Populations by race, sex, or age group
  - Time periods – secular trends
  - Crude vs age-adjusted mortality

SCTR Oklahoma Shared Clinical & Translational Resources

39

Mortality rates - example

	Community A			Community B		
Age	No. of Deaths	Mid-year Pop.	Rate	No. of Deaths	Mid-year Pop.	Rate
Young	1	1000	1	10	5000	2
Middle	15	3000	5	40	4000	10
Old	50	5000	10	20	1000	20
Total	66	9000	7.3	70	10000	7.0

SCTR Oklahoma Shared Clinical & Translational Resources

40

Mortality rates - example

	Community A			Community B		
Age	No. of Deaths	Mid-year Pop.	Rate per 1000	No. of Deaths	Mid-year Pop.	Rate per 1000
Young	1	1000	1	10	5000	2
Middle	15	3000	5	40	4000	10
Old	50	5000	10	20	1000	20
Total	66	9000	7.3	70	10000	7.0

SCTR Oklahoma Shared Clinical & Translational Resources

41

Activity 2

In groups of 2-3, using the epidemiologic concepts we have discussed, describe death rate patterns from the "unusual event".

- how do death rates vary between men and women overall?
- among the different socioeconomic classes?
- among men and women in different socioeconomic classes?
- among adults and children in different socioeconomic classes?

Can you guess what type of situation might result in such death rate patterns?

SCTR Oklahoma Shared Clinical & Translational Resources

42

Table 1.1 Deaths and Death Rates for an Unusual Event, by Sex and Socioeconomic Status

		Socioeconomic Status			
Sex	Measure	High	Middle	Low	Total
Males	Persons at risk	179	173	499	851
	Deaths	120	148	441	709
	Death rate (%)	67.0	85.5	88.4	83.3
Females	Persons at risk	143	107	212	462
	Deaths	9	13	132	154
	Death rate (%)	6.3	12.6	62.3	33.3
Both sexes	Persons at risk	322	280	711	1313
	Deaths	129	161	573	863
	Death rate (%)	40.1	57.5	80.6	65.7

Oklahoma Shared Clinical & Translational Resources

43

Table 1.1 Deaths and Death Rates for an Unusual Event, by Sex and Socioeconomic Status

		Socioeconomic Status			
Sex	Measure	High	Middle	Low	Total
Males	Persons at risk	179	173	499	851
	Deaths	120	148	441	709
	Death rate (%)	67.0	85.5	88.4	83.3
Females	Persons at risk	143	107	212	462
	Deaths	9	13	132	154
	Death rate (%)	6.3	12.6	62.3	33.3
Both sexes	Persons at risk	322	280	711	1313
	Deaths	129	161	573	863
	Death rate (%)	40.1	57.5	80.6	65.7

Oklahoma Shared Clinical & Translational Resources

44

Table 1.1 Deaths and Death Rates for an Unusual Event, by Sex and Socioeconomic Status

		Socioeconomic Status			
Sex	Measure	High	Middle	Low	Total
Males	Persons at risk	179	173	499	851
	Deaths	120	148	441	709
	Death rate (%)	67.0	85.5	88.4	83.3
Females	Persons at risk	143	107	212	462
	Deaths	9	13	132	154
	Death rate (%)	6.3	12.6	62.3	33.3
Both sexes	Persons at risk	322	280	711	1313
	Deaths	129	161	573	863
	Death rate (%)	40.1	57.5	80.6	65.7

Oklahoma Shared Clinical & Translational Resources

45

Table 1.2 Deaths and Death Rates for an Unusual Event, by Age and Socioeconomic Status

		Socioeconomic Status		
Age Group	Measure	High/Middle	Low	Total
Adults	Persons at risk	566	664	1230
	Deaths	287	545	832
	Death rate (%)	50.7	82.1	67.6
Children	Persons at risk	36	47	83
	Deaths	3	28	31
	Death rate (%)	8.3	59.6	37.3
All Ages	Persons at risk	602	711	1313
	Deaths	290	573	863
	Death rate (%)	48.2	80.6	65.7

Oklahoma Shared Clinical & Translational Resources

46

Table 1.2 Deaths and Death Rates for an Unusual Event, by Age and Socioeconomic Status

		Socioeconomic Status		
Age Group	Measure	High/Middle	Low	Total
Adults	Persons at risk	566	664	1230
	Deaths	287	545	832
	Death rate (%)	50.7	82.1	67.6
Children	Persons at risk	36	47	83
	Deaths	3	28	31
	Death rate (%)	8.3	59.6	37.3
All Ages	Persons at risk	602	711	1313
	Deaths	290	573	863
	Death rate (%)	48.2	80.6	65.7

Oklahoma Shared Clinical & Translational Resources

47

Table 1.2 Deaths and Death Rates for an Unusual Event, by Age and Socioeconomic Status

		Socioeconomic Status		
Age Group	Measure	High/Middle	Low	Total
Adults	Persons at risk	566	664	1230
	Deaths	287	545	832
	Death rate (%)	50.7	82.1	67.6
Children	Persons at risk	36	47	83
	Deaths	3	28	31
	Death rate (%)	8.3	59.6	37.3
All Ages	Persons at risk	602	711	1313
	Deaths	290	573	863
	Death rate (%)	48.2	80.6	65.7

Oklahoma Shared Clinical & Translational Resources

48

Age Adjustment
AKA Age Standardization

- Allows for better comparison of populations with different age profiles or distributions
- Important for health issues that disproportionately affects different age groups (cardiovascular disease, leukemia, car accidents)

SCTR Oklahoma Shared Clinical & Translational Resources

49

Case Fatality Rate

Case fatality rate = $\frac{\text{(\# of deaths due to certain cause)}}{\text{(\# of people } \textbf{affected by same cause}\text{)}}$

SCTR Oklahoma Shared Clinical & Translational Resources

50

Case Fatality Rate

- **Swine flu (H1N1)**
  - As of August 6, 2009: 1,462 deaths from 177,457 cases
  - 1,462/177,456 = 0.0082 or 0.82%
- **Breast cancer**
  - In one study, from 2000 to 2004, 71 deaths from 244 cases
  - 71/244 = 0.2910 or 29.10%

SCTR Oklahoma Shared Clinical & Translational Resources

51

Years of Potential Life Lost

- Years of potential life lost (YPLL)
  - Lost years of life before "standard" age of death
- Reference age=75
  - Age at death= 60 ➡ YPLL(75)=75-60=15
  - Age at death= 6 months ➡ YPLL(75)=75-.5=74.5
  - Age at death=80 ➡ YPLL(75)=0
- There may be different reference ages for different populations

SCTR Oklahoma Shared Clinical & Translational Resources

52

Ratio Measures

- Odds ratio
- Risk Ratio
- Rate Ratio

	Sick/Outcome Occurred/Effect Present	Not Sick/Outcome Did Not Occur/No Effect Present
Exposed/Risk Factor Present/Intervention	a	b
Not Exposed/Risk Factor Not Present/Control	c	d

SCTR Oklahoma Shared Clinical & Translational Resources

53

Ratio Measures

a = number of persons exposed and with disease

	Sick/Outcome Occurred/Effect Present	Not Sick/Outcome Did Not Occur/No Effect Present
Exposed/Risk Factor Present/Intervention	a	b
Not Exposed/Risk Factor Not Present/Control	c	d

SCTR Oklahoma Shared Clinical & Translational Resources

54

Ratio Measures

b = number of persons exposed but without disease

	Sick/Outcome Occurred/Effect Present	Not Sick/Outcome Did Not Occur/No Effect Present
Exposed/Risk Factor Present/Intervention	a	b
Not Exposed/Risk Factor Not Present/Control	c	d

SCTR Oklahoma Shared Clinical & Translational Resources

55

Ratio Measures

c = number of persons unexposed but with disease

	Sick/Outcome Occurred/Effect Present	Not Sick/Outcome Did Not Occur/No Effect Present
Exposed/Risk Factor Present/Intervention	a	b
Not Exposed/Risk Factor Not Present/Control	c	d

SCTR Oklahoma Shared Clinical & Translational Resources

56

Ratio Measures

d = number of persons unexposed and without disease

	Sick/Outcome Occurred/Effect Present	Not Sick/Outcome Did Not Occur/No Effect Present
Exposed/Risk Factor Present/Intervention	a	b
Not Exposed/Risk Factor Not Present/Control	c	d

SCTR Oklahoma Shared Clinical & Translational Resources

57

Ratio Measures

a + c = total number of persons with disease

	Sick/Outcome Occurred/Effect Present	Not Sick/Outcome Did Not Occur/No Effect Present
Exposed/Risk Factor Present/Intervention	a	b
Not Exposed/Risk Factor Not Present/Control	c	d

SCTR Oklahoma Shared Clinical & Translational Resources

58

Ratio Measures

b + d = total number of persons without disease

	Sick/Outcome Occurred/Effect Present	Not Sick/Outcome Did Not Occur/No Effect Present
Exposed/Risk Factor Present/Intervention	a	b
Not Exposed/Risk Factor Not Present/Control	c	d

SCTR Oklahoma Shared Clinical & Translational Resources

59

Risk ratio

$$\text{Risk ratio} = \frac{\text{Risk of disease in exposed group}}{\text{Risk of disease in comparison group}}$$

SCTR Oklahoma Shared Clinical & Translational Resources

60

Risk ratio

$$\text{Risk ratio} = \frac{a/a+b}{c/c+d}$$

SCTR Oklahoma Shared Clinical & Translational Resources

61

Example case

In an outbreak of varicella (chickenpox) in Oregon in 2002, varicella was diagnosed in 18 of 152 vaccinated children compared with 3 of 7 unvaccinated children.

SCTR Oklahoma Shared Clinical & Translational Resources

62

Example case

	Chickenpox	No chickenpox
Vaccinated	a = 18	b = 134
Not vaccinated	c = 3	d = 4

SCTR Oklahoma Shared Clinical & Translational Resources

63

Example case

Risk of varicella among vaccinated children = 18 / 152 = 0.118 = 11.8%

Risk of varicella among unvaccinated children = 3 / 7 = 0.429 = 42.9%

Risk ratio = 0.118 / 0.429 = 0.28

	Chickenpox	No chickenpox
Vaccinated	a = 18	b = 134
Not vaccinated	c = 3	d = 4

SCTR Oklahoma Shared Clinical & Translational Resources

64

Example case

The risk ratio is less than 1.0, indicating a decreased risk or protective effect for the exposed (vaccinated) children. The risk ratio of 0.28 indicates that vaccinated children were only approximately one-fourth as likely (28%, actually) to develop varicella as were unvaccinated children.

SCTR Oklahoma Shared Clinical & Translational Resources

65

Interpreting Ratio Measures

- =1 ➡ No difference in risk/odds/rate between two groups
- >1 ➡ Higher risk/odds/rate in one group compared to another
- <1 ➡ Lower risk/odds/rate in one group compared to another

SCTR Oklahoma Shared Clinical & Translational Resources

66

ACTIVITY 3

SCTR Oklahoma Shared Clinical & Translational Resources

67

		Disease (HepA)	
		Y	N
Exposure (salsa)	Y	470	195
	N	30	305

SCTR Oklahoma Shared Clinical & Translational Resources

68

INTERPRETING DATA

SCTR Oklahoma Shared Clinical & Translational Resources

Interpreting Data

- Is the information what you expected?
- Are there differences over time?
- Is the measure bigger or smaller than:
  - another area
  - another group of people
  - another situation/circumstance

Spend time with and get to know your data!!!

SCTR Oklahoma Shared Clinical & Translational Resources

Interpreting Data

- What if something happened due to chance?
  - P-value
- How precise are your results?
  - Confidence interval

SCTR Oklahoma Shared Clinical & Translational Resources

71

P-value

- The probability of getting some test statistic by "chance"
- The smaller the p-value, the more significant your results because the smaller the probability that it can be observed by chance
- Is affected by size of sample

SCTR Oklahoma Shared Clinical & Translational Resources

72

Confidence Interval

- An interval estimate of some statistical measure
- Usually set at 95%
- Can be interpreted as: if this experiment was repeated, you would get results within a specific interval 95 times out of 100 times
- Is affected by size of sample
- For ORs and RRs, confidence intervals including 1 (for example, 0.95-1.30) do not represent a difference

SCTR Oklahoma Shared Clinical & Translational Resources

73

Characterizations of Good Data

- Accuracy
  - "Closeness of arrow to bull's eye"
- Validity
  - Are the data for measures you're interested in?
- Reliability
  - Information has stable and consistent collection processes
- Timeliness
  - Less than 5 years old is a good standard
- Relevance
  - Are you using them for the right purposes?
- Completeness
  - Quality measure

SCTR Oklahoma Shared Clinical & Translational Resources

74

Can the data be trusted?

- Validity!!
- Research/study agenda
- Sample size and sampling methods
- P-value
- Confidence interval
- Associations of research/publisher

SCTR Oklahoma Shared Clinical & Translational Resources

75

What is Validity?

- Internal (your study population)
  - The ability to really measure what is supposed to be measured
  - The degree to which a measurement (of study) reaches a correct conclusion and represents the truth
- External (the world)
  - Generalizability of results

SCTR Oklahoma Shared Clinical & Translational Resources

76

Internal Validity

- Any systematic error that results in an incorrect estimate (invalidity) of your population
- All research, evaluation, or study has bias
- Less bias ➡ Better representation of the truth!!
- 3 classes or types
  - Selection bias
  - Observation/Information bias
  - Confounding

SCTR Oklahoma Shared Clinical & Translational Resources

77

Threats to External Validity

- Can we generalize our study findings to those not included in the study?

SCTR Oklahoma Shared Clinical & Translational Resources

Gordis, 2004

78

Association v. Causation

- Association is observed
  - Correlation, relationship, statistical dependence
  - Types of associations
    - Artifactual/spurious, non-causal, direct causal, non-direct causal
- Causation is inferred

SCTR Oklahoma Shared Clinical & Translational Resources

Gordis, 2004

79

Comprehend and Scrutinize

- Study population (who and how many?)
- Study methodology
  - Sampling
  - Analysis
- Author disclosure
  - Health statistics are ALWAYS influenced by an individual's/organization's biases
- Relevancy and coherence of findings

SCTR Oklahoma Shared Clinical & Translational Resources

80

ORGANIZING DATA

SCTR Oklahoma Shared Clinical & Translational Resources

Organizing data

A **variable** can be any characteristic that differs from person to person

- Height
- Weight
- Gender
- Vaccine status

The **value** of a variable is the number or descriptor that applies to a particular person

SCTR Oklahoma Shared Clinical & Translational Resources

ID	Date of Diagnosis	Town	Age (Years)	Sex	Hosp	Jaundice	Outbreak	IV Drugs	IgM Pos
01	01/05	B	74	M	Y	N	N	N	Y
02	01/06	J	29	M	N	Y	N	Y	Y
03	01/08	K	37	M	Y	Y	N	N	Y
04	01/19	J	3	F	N	N	N	N	Y
05	01/30	C	39	M	N	Y	N	N	Y
06	02/02	D	23	M	Y	Y	N	Y	Y
07	02/03	F	19	M	Y	Y	N	N	Y
08	02/05	I	44	M	N	Y	N	N	Y
09	02/19	G	28	M	Y	N	N	Y	Y
10	02/22	E	29	F	N	Y	Y	N	Y

SCTR Oklahoma Shared Clinical & Translational Resources

83

ID	Date of Diagnosis	Town	Age (Years)	Sex	Hosp	Jaundice	Outbreak	IV Drugs	IgM Pos
01	01/05	B	74	M	Y	N	N	N	Y
02	01/06	J	29	M	N	Y	N	Y	Y
03	01/08	K	37	M	Y	Y	N	N	Y
04	01/19	J							Y
05	01/30	C							Y
06	02/02	D							Y
07	02/03	F	19	M	Y	Y	N	N	Y
08	02/05	I	44	M	N	Y	N	N	Y
09	02/19	G	28	M	Y	N	N	Y	Y
10	02/22	E	29	F	N	Y	Y	N	Y

Typically, each row is called a **record** or an **observation** and represents one person or case of disease

SCTR Oklahoma Shared Clinical & Translational Resources

84

ID	Date of Diagnosis	Town	Age (Years)	Sex	Hosp	Jaundice	Outbreak	IV Drugs	IgM Pos
01	01/05	B	74	M	Y	N	N	N	Y
02	01/06	J	29	M	N	Y	N	Y	Y
03	01/08	K	37	M	Y	Y	N	N	Y
04	01/19							N	Y
05	01/30							N	Y
06	02/02	D	23	M	Y	Y	N	Y	Y
07	02/03	F	19	M	Y	Y	N	N	Y
08	02/05	I	44	M	N	Y	N	N	Y
09	02/19	G	28	M	Y	N	N	Y	Y
10	02/22	E	29	F	N	Y	Y	N	Y

Each column is called a **variable** and contains information about one characteristic

SCTR Oklahoma Shared Clinical & Translational Resources

85

Organizing data

- Code values to numeric for ease of analysis
  - Gender: male = 1, female = 2
  - Town: Ada = 1, Holdenville = 2, Ardmore = 3, Sulphur = 4
- HOWEVER do not transform data from collection instrument to database/spreadsheet
- Create a codebook to document!

SCTR Oklahoma Shared Clinical & Translational Resources

86

ACTIVITY 4

SCTR Oklahoma Shared Clinical & Translational Resources

DISPLAYING DATA

SCTR Oklahoma Shared Clinical & Translational Resources

Displaying Data

- Important to disseminate/present research findings
- Visual aids can provide clear synthesis of findings
- Display data in tables or charts/graphs
- Data display should be self-explanatory and stand alone

SCTR Oklahoma Shared Clinical & Translational Resources

89

Choosing a Table or Chart

- Amount of data to be presented
- Level of detail needed
- Relationship within the data you wish to communicate
- Type of presentation
- Type of audience

SCTR Oklahoma Shared Clinical & Translational Resources

90

Tables

- Clear and simple summary using rows and columns
- Use to show many and precise numerical values in small space
- Primarily in manuscripts, technical reports, posters
- Often difficult to read in PowerPoint presentations

SCTR Oklahoma Shared Clinical & Translational Resources

91

Guidelines for Tables

- Rows and columns should have clear labels
- No unnecessary digits
- Lines or spaces separate items
- Ensure sufficient space between columns and rows
- Avoid messy or crowded tables – consider dividing into 2

SCTR Oklahoma Shared Clinical & Translational Resources

92

Table Titles

- Concisely describe purpose or content of table
- Should be clear and informative
- Descriptive of the data
  - Measures
  - Time frame
  - Place
  - Sample size (when appropriate)

SCTR Oklahoma Shared Clinical & Translational Resources

93

Table Footer

- Use asterisks or other symbols to highlight important information about variables or values in table and describe in footer
- E.g. significance levels, covariates in model, data source

SCTR Oklahoma Shared Clinical & Translational Resources

94

Table 1 Example

Table 1. Demographic Characteristics of a Sample of Employed Workers From the National Health Interview Survey ($n = 89,366$), 2004–2009[a]

Characteristic	No. of Subjects Interviewed	Weighted %
Gender		
Men	44,473	50.8
Women	44,893	49.2
Highest level of education		
Did not complete high school	11,088	9.8
High school graduate	23,285	25.8
Some college	27,611	31.9
Bachelor's degree	17,865	21.6
Master's degree	6,664	8.0
Professional school degree or doctoral degree	2,405	2.9
Race/ethnicity		
Hispanic	16,893	11.9
Non-Hispanic white	54,339	71.3
Non-Hispanic black	13,133	12.1
Non-Hispanic other	5,001	4.7

[a] Numbers may not sum to total because of missing data.

95

Charts/Graphs

96

http://ieatalp.weebly.com/

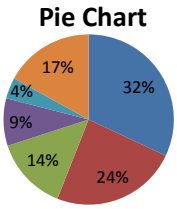
Charts/Graphs

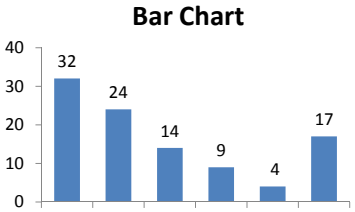
- Uses bars, lines, or dots to visually represent data
- Use to show trends, patterns, and relationships across data set
- Use when general pattern is more important than exact data values
- Useful in oral and poster presentations, for non-technical audience but also manuscripts and technical reports

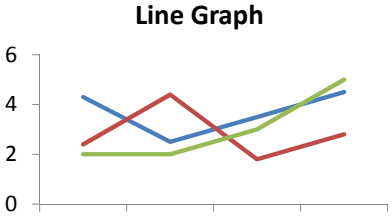
SCTR Oklahoma Shared Clinical & Translational Resources

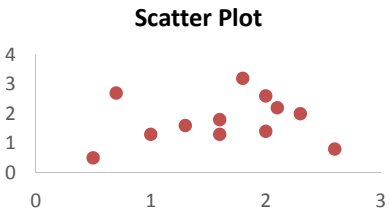
97

Types of Charts





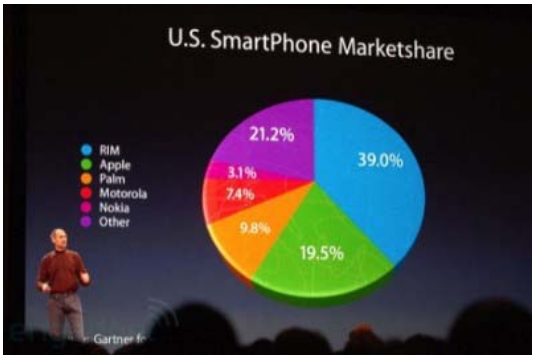




98

Pie Chart

- Displays parts to a whole
- Not easy for the audience to compare angles (heights and lengths of bars are easier)
- Can be deceiving

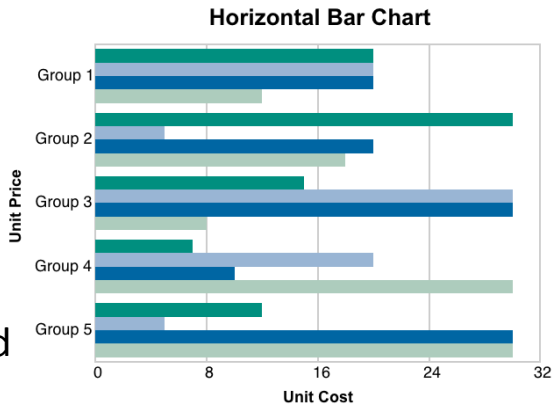


SCTR Oklahoma Shared Clinical & Translational Resources

http://www.engadget.com/2008/01/15/live-from-macworld-2008-steve-jobs-keynote/

Bar Chart

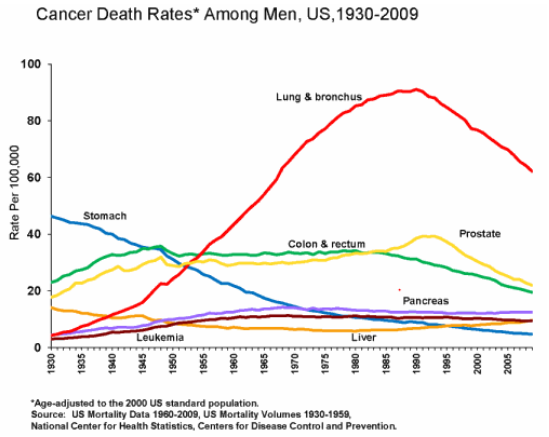
- Uses horizontal or vertical bars to compare quantities of categorical data
- Bars do not touch
- Histograms are similar, but for continuous data and bars touch



SCTR Oklahoma Shared Clinical & Translational Resources

Line Graph

- Series of data points connected by straight lines
- Used to show change over some type of order (usually time)



SCTR Oklahoma Shared Clinical & Translational Resources

http://www.cancer.org/research/cancerfactsstatistics/cancerfactsfigures2013/cancer-statistics-2013-slide-presentation.pdf

101

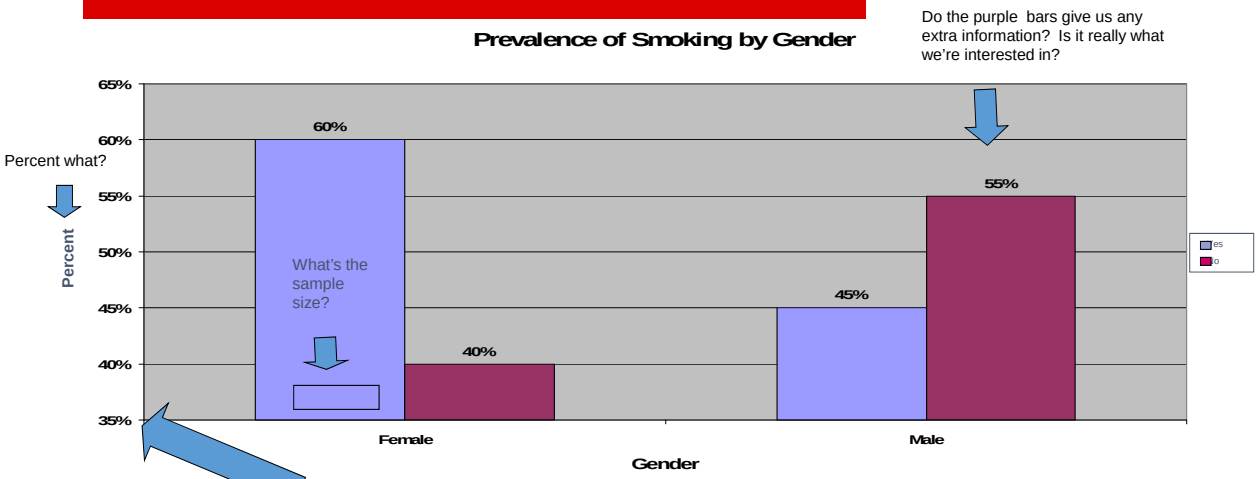
Guidelines for Charts/Graphs

- Clear and concise title
- Labels on x and y axis with variable and units of measure
- Scale should start at zero and be clearly marked
- Any break in scale should be labeled
- Avoid clutter and 3-D effects
- Use legend or proper labels on bars and lines

SCTR Oklahoma Shared Clinical & Translational Resources

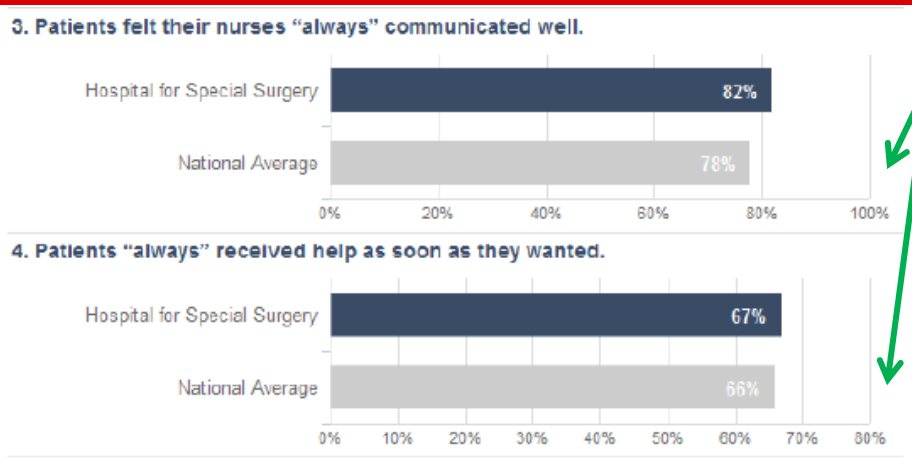
102

What changes do you suggest?



Y-axis doesn't start at zero. This distorts the graph and makes it seems that there are more than double the percent of female smokers than male smokers.

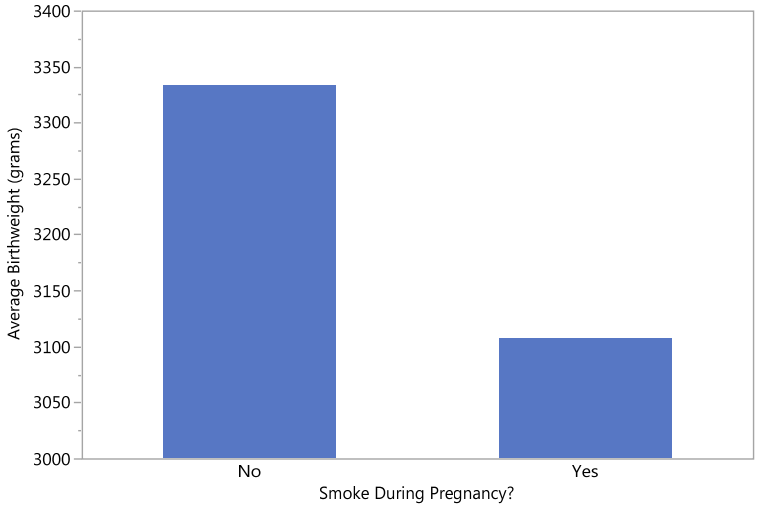
What changes do you suggest?



The scale differs between panels – level of satisfaction within a survey is lower for Question 4.

Citation: Robbins, N. 2014. Misleading Healthcare Graph. Forbes. https://www.forbes.com/sites/naomirobbins/2014/01/08/misleading-healthcare-graph/#438f10b76a17

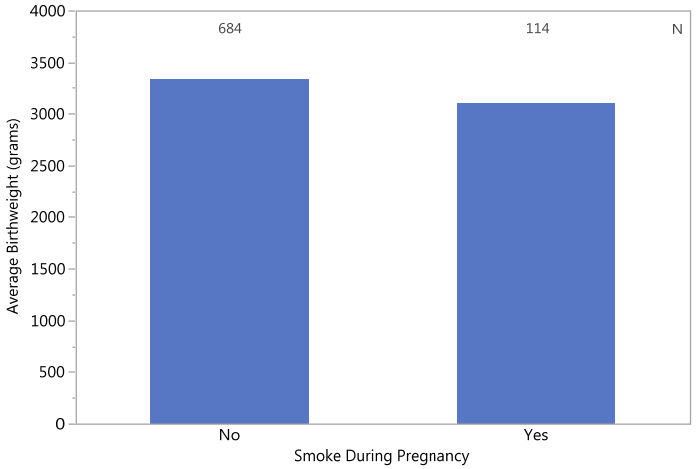
North Carolina Births – smoking by birthweight



SCTR Oklahoma Shared Clinical & Translational Resources

105

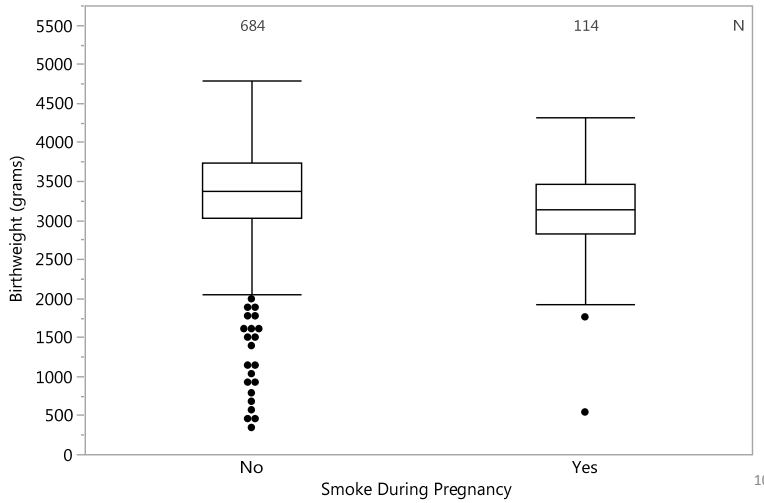
North Carolina Births – corrected



SCTR Oklahoma Shared Clinical & Translational Resources

106

North Carolina Births – alternative graph



SCTR Oklahoma Shared Clinical & Translational Resources

107

Data Displays

- Choose table or chart based on purpose, audience, and type of data
- Data display should stand alone (titles, labels, legends)
- Ensure display portrays data accurately and in a way that is clear to audience and not deceiving

SCTR Oklahoma Shared Clinical & Translational Resources

108

Future topics

- Principles of data capture, processing/coding, and cleaning
  - Now that you have data, what are the first steps in preparing for the analysis?
  - Important steps in data cleaning – pitfalls and problems to avoid
- Accessing and analyzing publicly-available data
  - Lab/applied hands-on session
  - Data Into Action training
- Principles of statistical analysis
  - Relate to the type of data and type of research design
  - Stress interpretation
- Others?

SCTR Oklahoma Shared Clinical & Translational Resources

109

Thank you for having us!

Questions?

110