MEDICAL COLLEGE OF WISCONSIN

Analyzing Discrete Data

Kwang Woo Ahn, PhD

Sponsored by the Clinical and Translational Science Institute (CTSI)
and the Department of Population Health / Division of Biostatistics

Speaker disclosure

In accordance with the ACCME policy on speaker disclosure, the speaker and planners who are in a position to control the educational activity of this program were asked to disclose all relevant financial relationships with any commercial interest to the audience. The speaker and program planners have no relationships to disclose.

Outline

- Types of data
- Describing categorical data
- Graphical displays of categorical data
- Measure of association for binary outcomes
- Risk difference/Relative risk/Odds ratio
- Comparing two proportions
- Concluding remarks

Types of data

- Categorical data
  - Nominal (no ordering)
    - Race/ethnicity, Type of disease, Eye color
  - Ordinal (ordered)
    - Toxicity Grade, Disease stage
- Continuous Data
  - Systolic blood pressure
  - Age
  - White Blood Cell (WBC) count

Binary data

- Binary (Dichotomous) Data
  - Yes/No outcome
  - Gender (Male vs. Female)
  - Response vs. no response to treatment
  - Alive vs. Dead

Describing categorical data

- Frequencies: number of patients in a particular category
- Proportions or Relative frequencies: number of patients in a particular category divided by the total number of patients
  - Often expressed in percentages

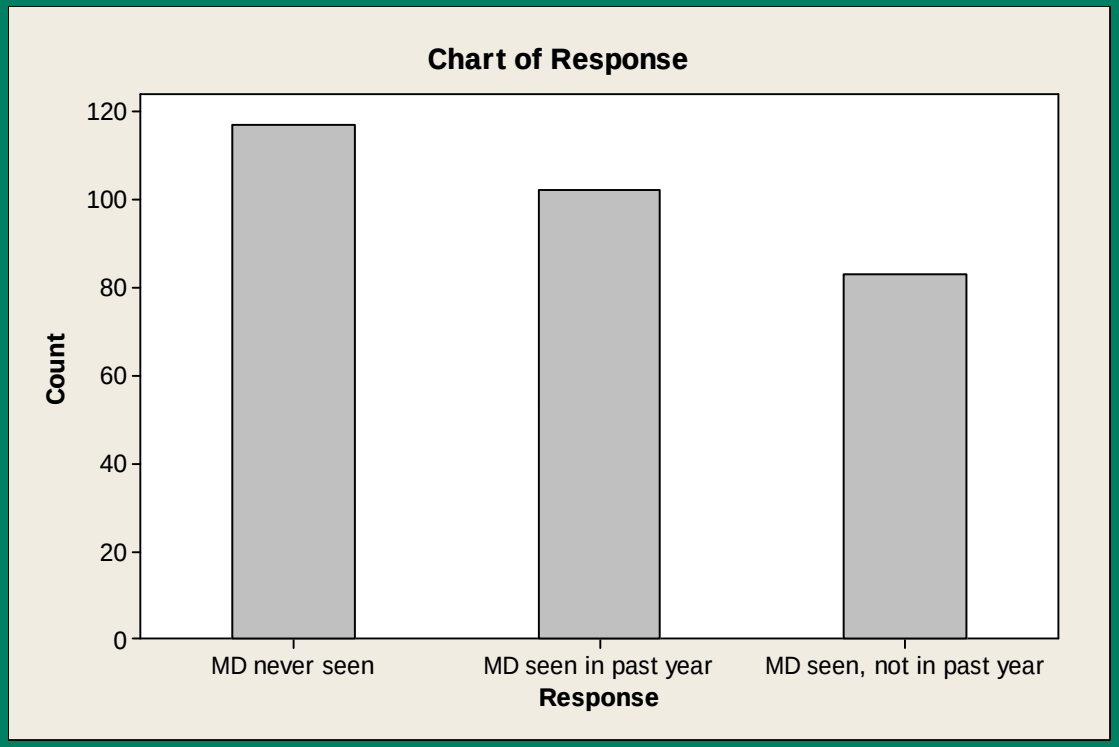
Example of frequency table

- Study of use of medical care by adults experiencing chest pain in the past year

Response	Frequency	Relative Frequency
MD seen in past year	102	0.34
MD seen, not in past year	83	0.27
MD never seen	117	0.39
Total	302	1

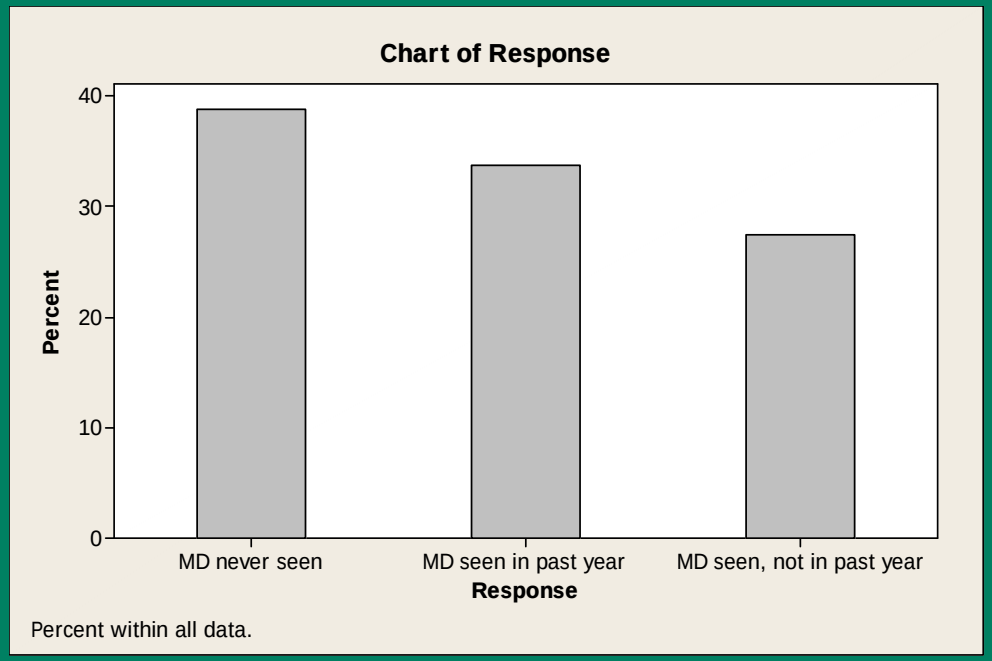
Graphical displays of categorical data

- Bar Chart: height of bars represent number of individuals in that category



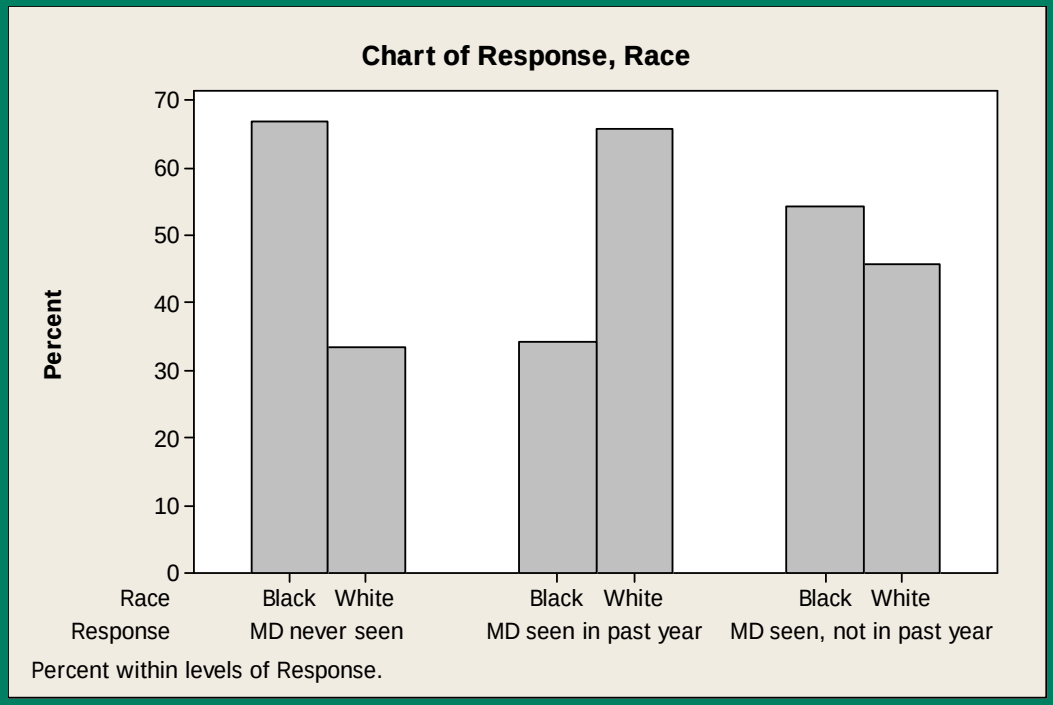
Graphical displays of categorical data

- Alternatively, the relative frequency or percentage can be plotted on the y axis



Graphical displays of categorical data

- Bar charts can be shown side by side to compare levels of another factor



Measures of association

- Risk difference: difference in event proportions
- Relative risk: ratio of event proportions
- Odds ratio: ratio of event odds

Example

- NEJM 318: 262-264, 1988
- Relationship between aspirin use and heart attacks
- Study Design: 5 year, blind and randomized study
- Physicians took one aspirin tablet or a placebo every other day
- Outcome: Fatal attack, nonfatal attack, and no attack

Example

		Outcome		
		Attack	No Attack	Total
Treatment	Placebo	189	10,845	11,034
	Aspirin	104	10,933	11,037
	Total	293	21,778	22,071

Example

- Aspirin:
  - Proportion with heart attack is 104/11037=0.009 or 0.9%
- Placebo:
  - Proportion with heart attack is 189/11034=0.017 or 1.7%
- Risk difference
  - Difference in sample proportions p_1-p_2
- Risk of heart attack for physicians with aspirin minus risk of heart attack for physicians with placebo
- 0.9%-1.7% = -0.8%

Interpretation

- If 1000 patients were given aspirin, this would reduce the number of heart attack by 8, compared to if all patients had not been given aspirin.

Relative Risk

- Also called Risk Ratio
- Ratio of proportions

$$RR = \frac{P_1}{P_2}$$

Example

- Relative risk of heart attack in the aspirin group compared to the placebo group is

$$\text{RR} = \frac{0.009}{0.017} = 0.53$$

- The aspirin group is 0.53 times as likely to result in heart attack as the placebo group.

Interpretation

- Patient's individual risk: If a patient were given an aspirin compared to the placebo group, their personal risk of heart attack would be reduced almost in half.

Comments on Relative Risk

- Directional Interpretation of RR:
  - RR=1: risk of heart attack is the same for both groups
  - RR>1: risk of heart attack is higher for the aspirin group than the placebo group
  - RR<1 risk of heart attack is lower for the aspirin group than the placebo group

Comments on Relative Risk

- Interpretation depends on “reference group” in denominator
  - RR of heart attack for placebo versus aspirin is

    RR*= 0.017/0.009=1.89 (=1/RR=1/0.53)
  - Reference group is typically “control” group or “unexposed” group
  - Interpretation: A patient receiving a placebo will have almost twice the risk of heart attack as he/she would have had an aspirin

Risk Difference vs. Relative Risk

- Risk difference gives the impact of the treatment in terms of the absolute risk of the outcome
  - More interpretable with regards to public health impact
- Relative risk gives the impact of one treatment relative to the other for an individual
  - No consideration of the absolute risk for that individual

Odds Ratio

- Odds ratio is the ratio of the odds of the event for one group divided by the odds of the event for the other group
- Definition of Odds

$$\text{Odds} = \frac{P}{1-P} = \frac{P(\text{Disease})}{P(\text{No Disease})}$$

Example of Odds

- Odds of heart attack for the aspirin group
  - Probability of heart attack=0.009

$$\text{Odds} = \frac{0.009}{1 - 0.009} = 0.0091$$

- Odds of heart attack for the placebo group
  - Probability of heart attack=0.017

$$\text{Odds} = \frac{0.017}{1 - 0.017} = 0.0173$$

Odds ratio

- Odds is not the risk, but is a *function* of the risk
- Risk can be compared between two groups using the relative risk
- Similarly, we can compare odds between two groups using the odds ratio

Odds ratio

- Odds ratio is the odds for group 1 divided by the odds for group 2

$$OR = \frac{Odds_1}{Odds_2} = \frac{p_1/(1-p_1)}{p_2/(1-p_2)}$$

Odds Ratio Example

- Odds ratio for heart attack in the aspirin group compared to the placebo group

$$\text{OR} = \frac{0.0091}{0.0173} = 0.526$$

- The odds of heart attack for the aspirin group is 0.526 times the odds of heart attack for the placebo group

Interpretation of OR

- The aspirin group is associated with an estimated 47.4% reduction (since OR=0.526) in the odds of heart attack compared to the placebo group.

Comments on Odds Ratio

- Directional Interpretation of OR:
  - OR=1: Odds of heart attack is the same for both groups
  - OR>1: Odds of heart attack is higher for the aspirin group than the placebo group
  - OR<1: Odds of heart attack is lower for the aspirin group than the placebo group

Comments on Odds Ratio

- Interpretation depends on “reference group” in denominator
  - OR of heart attack for the placebo group versus the aspirin group is

    OR*= 0.0173/0.0091=1.90 (=1/OR=1/0.526)
  - Reference group is typically “control” group or “unexposed” group
  - Interpretation: A patient receiving a placebo will have almost twice the odds of heart attack as he/she would have had an aspirin

OR vs. RR

- Same direction of association

$$OR < 1 \Leftrightarrow RR < 1$$

$$OR = 1 \Leftrightarrow RR = 1$$

$$OR > 1 \Leftrightarrow RR > 1$$

OR vs. RR

- OR is very close to RR when the prevalence of disease is small
- In this study the overall risk of heart attack was 1.3%
  - Relative Risk: 0.53
  - Odds Ratio: 0.526

OR vs. RR

- Relative risk is more interpretable than odds ratio
- Why bother with OR?
  - OR can be calculated in both cohort studies as well as case-control studies
  - RR can only be calculated in cohort studies

Comparing Two Proportions

- Chi-squared test and two-sample z-test are based on large samples
  - Cell counts are all at least 5
- Otherwise use Fisher's Exact test
  - No minimum sample size requirement
  - Gives exact, not approximate, p-value
  - Difficult calculations; computer necessary

Example

- Weindling et al. (BMJ, 1986)
- Comparison of eye vision health of juvenile delinquent boys and a control group.
- Outcome: Whether or not the boy wears glasses

Example

		Delinquent	Non-delinquent	Total
Outcome	Glasses	1	5	6
	No Glasses	8	2	10
	Total	9	7	16

Fisher's Exact test: results

- Two-sided p-value: approximately twice the one-sided p-value
  - P=0.035
- Reject the null hypothesis that the proportion of juvenile delinquents wearing glasses is the same as the proportion of non-juvenile delinquents wearing glasses

Concluding remarks

- Bar charts can be used for graphical displays
- Risk difference, relative risk, and odds ratio can be used for measure of association
- Fisher's exact test can be used for comparing two proportions with small sample sizes

Resources

- The Clinical and Translation Science Institute (CTSI) supports education, collaboration, and research in clinical and translational science: www.ctsi.mcw.edu
- The Biostatistics Consulting Service provides comprehensive statistical support

  http://www.mcw.edu/biostatsconsult.htm

Free drop-in consulting

- MCW/Froedtert/CHW:
  - Monday, Wednesday, Friday 1 – 3 PM @ CTSI Administrative offices (LL772A)
  - Tuesday, Thursday 1 – 3 PM @ Health Research Center, H2400
- VA: 1st and 3rd Monday, 8:30-11:30 am
  - VA Medical Center, Building 70, Room D-21
- Marquette: 2nd and 4th Monday, 8:30-11:30 am
  - Olin Engineering Building, Room 338D