Statistics in Clinical Practice

Jacob P. Gettig, PharmD, MPH, BCPS
Assistant Dean for Postgraduate Education
Associate Professor of Pharmacy Practice
Midwestern University Chicago College of Pharmacy

Disclosure

- Dr. Gettig has no conflicts of interest to disclose.

Learning Objectives

- At the end of this session, the learner will be able to:
  - Define and interpret the following: nominal data, ordinal data, continuous data, Type I error, Type II error, alpha, beta, power, p-values and confidence intervals.
  - Describe the factors that affect statistical power.
  - Compare, contrast and calculate absolute risk, absolute risk reduction/increase, relative risk, relative risk reduction/increase, odds ratio and number needed to treat/harm

Learning Objectives (cont.)

- At the end of this session, the learner will be able to:
  - Select the appropriate parametric or nonparametric statistical test given a set of variables and a hypothesis
  - Explain how statistical results in clinical studies can be used to make clinical decisions.

How confident are you with regard to your ability to interpret statistics?

A. Very confident
B. Moderately confident
C. Somewhat confident
D. Not at all confident
E. This session is about statistics? Where's the door?!

Basic Terms/Concepts

- Types of data
  - Nominal
  - Ordinal
  - Continuous
    - Interval
    - ratio
- Error and statistical significance terms
  - Alpha
  - Beta
  - Type I
  - Type II
  - Power
  - P-value
  - Confidence interval

Types of Data

- Nominal
  - Categorical
  - Examples: sex, race, yes/no
- Ordinal
  - Ranked/directional
  - Distance between points not absolute
  - Examples: pain scale, Likert scale
- Continuous
  - Quantitative
  - Distance between points absolute
  - Examples: weight, SCr, temperature

General Rule about Types of Data

- Higher level data can be transformed into lower level data, but not the converse.
- Example:
  - 4 patients
    - SBPs: 119mmHg, 127mmHg, 117mmHg, 136mmHg
    - Assume ≤ 120mmHg is "at goal"
    - How many patients are at goal?
    - If we started with how many patients were at goal, would you be able to determine their individual SBPs?

Table 1. Baseline Characteristics of the Participants

Characteristic	Sustained-Release Bupropion (n = 300)	Placebo (n = 300)
Age, mean (SD), y	44.0 (10.9)	44.4 (11.3)
Women, No. (%)	212 (70.7)	208 (69.3)
Married or living with a partner, No. (%)	117 (39.0)	113 (37.7)
Monthly family income <$1800, No. (%)	158 (52.6)	164 (54.6)
≤High school graduate, No. (%)	151 (50.3)	149 (49.7)
No. of cigarettes smoked per day, mean (SD)	16.1 (7.5)	17.1 (8.5)
Smoke mentholated cigarettes, No. (%)	235 (78.3)	236 (78.7)
Fagerström score, mean (SD)*	4.6 (2.1)	4.7 (1.9)
No. of previous serious attempts to quit, mean (SD)	2.1 (4.7)	2.2 (4.2)
Salivary cotinine, mean (SD), ng/mL	287.2 (138.8)	296.5 (147.0)
Exhaled carbon monoxide, mean (SD), ppm	22.1 (13.2)	23.3 (15.2)
Previous use of sus		24 (8.0)
Other smokers in th		96 (32.0)
Weight, mean (SD),		81.6 (20.1)
Body mass index, n		28.7 (6.3)
CES-D, mean (SD)†		11.9 (8.7)
Possible clinical dep		84 (28.0)

A = Nominal

B = Ordinal

C= Continuous

*The Fagerström test s… of nicotine dependence. …te greater levels

†Center for Epidemiologic Studies Depression Scale (CES-D) scores can range from 0 to 60. Scores of 16 or higher indicate the likelihood of clinical depression because it represented the 80th percentile in a representative population.[9]

Adapted from Ahluwalia JS, Harris KJ, Catley D, Okuyemi K, Mayo M. Sustained-Release Bupropion for Smoking Cessation in African Americans. *JAMA* 2002 Jul; 288(4):468-474.

Hypotheses

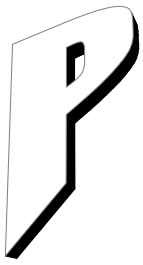
- Null hypothesis (H_o)
  - Assumes no difference between therapies
  - Goal of studies is usually to reject H_o
- Research/alternate hypothesis (H_A)
  - Assumes a difference between therapies
  - Goal of studies is usually to accept H_A

Type I Error

- Also called “alpha” error
- Null hypothesis is incorrectly rejected
- A difference is inferred; however, there is no true difference → false positive
- **alpha $\leq$ 0.05** is generally accepted

What does the ‘p’ value mean?

- Probability (p) is a numeric estimate of the likelihood of occurrence of an event
- The probability of any given event (A) will always range between 0 and 1.
  - p(A) = 0
  - p(A) = 1
- What is the p of randomly drawing a ‘7’ from ten ping-pong balls (0 - 9)?



What does the 'p' value mean?

- Every inferential statistical test has a 'test statistic' (t, F, χ^2) and a probability of that result (p)
- 'p' is the probability that there is no difference, no effect, or no relationship between the groups in the entire population of interest
  - Probability that the **null hypothesis** is **true**
  - How likely is the observed difference, effect, or relationship due to random chance?

What does the 'p' value mean?

- The p value is used in determining whether or not the null hypothesis should be accepted or rejected
- If 'p' is less than or equal to alpha, the null hypothesis must be rejected
  - What about the research hypothesis?
- If 'p' is greater than alpha, the null hypothesis must be accepted (or not rejected)
  - What about the research hypothesis?

Please complete question #1 on worksheet.

Confidence Intervals

- Estimates the range of values likely to contain the **true** value for a population
  - Most population values (mean, %) are practically impossible to obtain
- The width of a confidence interval depends on
  - The amount of variability in the sample data
  - The degree of confidence the researchers wish to have that their interval contains the true value

Confidence Intervals

- Reported with means, percentages, relative risk/hazard ratio, and odds ratio
  - Can be reported for almost anything
- **Do not** interpret the CI as the percentage of the population that is distributed within the range!
- 95% CI is standard

Type II Error

- Also called "beta" error
- Null hypothesis is incorrectly accepted (or not rejected)
- No difference is detected; however, there is a true difference → false negative
- **beta ≤ 0.20** is generally accepted
- Chance for Type II error generally decreases as sample size increases

Can Type I and Type II error occur simultaneously for the same endpoint in a study?

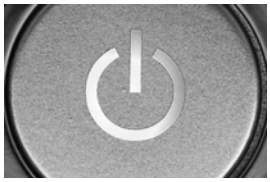
A. Yes
B. No

Type I/II Error

		REALITY	
		"Better" than placebo	No difference from placebo
STUDY CONCLUSIONS	"Better" than placebo		Type I error (α)
	No difference from placebo	Type II error (β)	

Statistical Power

- Power = 1 – beta
- Refers to the statistical test's ability to detect a true difference
- **power ≥ 0.8** is generally accepted



80% POWER!!!

Factors Affecting Power

$$\text{Power} \propto \frac{\text{Sample Size, Effect Size, Alpha}}{\text{Std. Dev..}}$$

- Sample size and power are directly related.
- Effect size and power are directly related.
- Alpha and power are directly related.
- Standard deviation and power are inversely related.

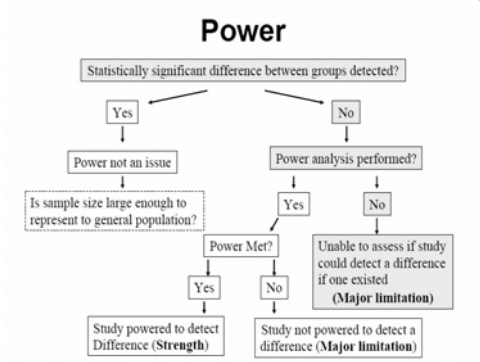
A researcher seeks to determine whether there is a significant difference in weight loss between patients who take Drug A and patients who take Drug B

- Null hypothesis: There is no difference in weight loss between groups.
- The researcher wants to detect a difference of 10% weight loss in the 2 groups.
- Upon review of previous literature, the researcher estimates the standard deviation at 8 pounds.
- Alpha is set at 0.05
- Sample size to achieve 80% power is 250 patients per group.

A = power increases, B = power decreases, C = no effect on power

- What if…
  - The researcher changed his mind and wanted to wants to detect a difference of 3% weight loss between groups?
  - The researcher underestimated the std. dev? It is actually 15 lbs.
  - The researcher wants to set a more conservative alpha at 0.01?
  - The researcher could only recruit 100 patients per group?

Power



Adapted from: Bryant PJ. A Practical Evidence Based Medicine Process for the Clinician. *ASHP Midyear Meeting*, 2006.

MEASURES OF RISK

AR, ARR, ARI
RR, OR, RRR, RRI
NNT, NNH

Caveats to Measures of Risk

- Can on be calculated if the outcome measure is dichotomous
  - **Beneficial effect (Y/N)**
  - **ADR or harmful effect (Y/N)**
  - **Event occurred or did not occur**
- Could not calculate these from continuous outcome measures:
  - **BP**
  - **LDL**
  - **Serum potassium**
- However, if continuous outcomes measures were transformed to dichotomous (nominal), one could calculate these:
  - **Met BP goal, did not meet BP goal**
  - **LDL ≤100 mg/dL, LDL > 100 mg/dL**
  - **Serum potassium WNL, serum potassium not WNL**

Caveats to Measures of Risk

- Always keep results in perspective
  - Are you referring to patients who received a medication or did not receive a medication?
  - Are you referring to a positive (desired) outcome or a negative (adverse) outcome?
  - In the case of head-to-head (comparative trials), which medication is the comparator and which is the intervention?
- By appropriately orienting yourself to the results, you will avoid misinterpretation of them.

Representing Risk

- Many epidemiologists (or clinicians) use 2 x 2 tables to visualize results.
- These tables show:
  - How many were exposed or not exposed to the risk factor in question
    - In the case of pharmacoepidemiology, the risk factor is usually a medication.
  - How many had the outcome of interest or did not have the outcome of interest
    - This can be a positive or negative outcome.
- Data not typically laid out in 2x2 tables in clinical trial manuscripts

Representing Risk

		Outcome		
		Y	N	
Exposure	Y	A	B	M1
	N	C	D	M2
		N1	N2	N

A = # of exposed persons with outcome
B= # of exposed persons without outcome
C= # of non-exposed persons with outcome
D= # of non-exposed persons without outcome

M1 = total study sample exposed
M2 = total study sample not exposed
N1 = total study sample with outcome
N2 = total study sample without outcome
N= total study sample

Layout in Clinical Trials

Table 2. **Incidence and Relative Risk of Confirmed Cardiovascular End Points.**

End Point	Aspirin (N=19,934)	Placebo (N=19,942)	Relative Risk (95% CI)*	P Value
	no. of events			
Major cardiovascular event†	477	522	0.91 (0.80–1.03)	0.13
Stroke	221	266	0.83 (0.69–0.99)	0.04
Ischemic	170	221	0.76 (0.63–0.93)	0.009
Hemorrhagic	51	41	1.24 (0.82–1.87)	0.31
Fatal	23	22	1.04 (0.58–1.86)	0.90
Nonfatal	198	244	0.81 (0.67–0.97)	0.02
Myocardial infarction	198	193	1.02 (0.84–1.25)	0.83
Fatal	14	12	1.16 (0.54–2.51)	0.70
Nonfatal	184	181	1.01 (0.83–1.24)	0.90
Death from cardiovascular causes	120	126	0.95 (0.74–1.22)	0.68
Transient ischemic attack	186	238	0.78 (0.64–0.94)	0.01
Coronary revascularization	389	374	1.04 (0.90–1.20)	0.61
Death from any cause	609	642	0.95 (0.85–1.06)	0.32

"Yes" exposure

"No" exposure

"Yes" outcomes

* CI denotes confidence interval.
† A major cardiovascular event was defined as a nonfatal myocardial infarction, a nonfatal stroke, or death from cardiovascular causes.

Adapted from: N Engl J Med 2005;352:1293-304.

What is Absolute Risk?

- Absolute risk is the risk or rate of an event in a defined period of time.
  - Sometimes called incidence or incidence rate
- Helps readers make sense of the true risks and benefits of treatment
  - Relative risk & odds ratio do not help the reader consider how common an outcome is to start with.
- Using the previous 2 slides as examples:
  - The AR for the outcome would be A/M_1 for those exposed (intervention group)
  - The AR for stroke in the aspirin group would be 221/19,934 = 0.011 = **1.1%**

What is ARR?

- Absolute Risk Reduction (ARR)
  - Difference in absolute risks of an outcome likely to prevent a patient from experiencing harm or experiencing a beneficial outcome
  - Related to Number Needed to Treat (NNT)
- Using the previous 2 slides as examples:
  - $ARR = A/M_1 – C/M_2$
  - The ARR for preventing stroke while taking ASA = $AR(stroke)_{ASA} – AR(stroke)_{placebo}$ = $|[(221/19934) – (266/19942)]| = |0.011 – 0.013| = |-0.002| =$ **0.2%**

What is ARI?

- Absolute Risk Increase (ARI)
  - Difference in absolute risks of an outcome likely to result in harm to a patient
  - Related to Number Needed to Harm (NNH)
- Calculated the same way as ARR
- Keep ARs in perspective to determine whether result is an ARR or ARI

What is NNT/NNH?

- Number needed to treat (NNT) tells you how many patients would need to be treated with the intervention before 1 patient would be prevented harm or experience a benefit.
  - NNT = 1/ARR
  - Want NNTs to be small
- Number needed to harm (NNH) tells you how many patients would need to be treated with the intervention before 1 patient would experience harm.
  - NNH = 1/ARI
  - Want NNHs to be large

What is NNT?

- For the previous example:
  - 1/ARR = 1/0.002 = 500
    - If you do not want to go back and forth between decimal and percent, you can divide 100 by ARR in its percent form (i.e., NNT = 100/0.2 = 500)
- Therefore, for every 500 patients that receive ASA instead of placebo (i.e., no therapy), 1 will be spared a stroke.

What is NNT/NNH?

- In clinical practice, NNTs and NNHs can be compared to weigh risks versus benefits.
- Should take into account how severe the outcomes are in question.
  - What's a more clinically important beneficial outcome?
    - Getting to goal BP or preventing stroke?
  - What's a more clinically important adverse outcome?
    - Experiencing a headache or an embolism?

Please complete questions #2a – 2d and #3a – 3d on worksheet.

Odds ratios vs. Relative Risks

- 4 marbles: 1 is red; 3 are blue
  - odds of choosing a red marble are 3 to 1 against choosing a red marble = 0.33 = 1/3
  - probability of choosing a red marble is 0.25 or 1 in 4
- Odds ratios (ORs) are often used in case-control studies and in regression models.
- Relative risks (RRs) are calculated in other study designs.
- Hazard ratios (HRs) can be interpreted similar to RRs.

		Outcome		
		Y	N	
Exposure	Y	A	B	M1
		C	D	M2
	N			
		N1	N2	N

Odds ratio = (A/B) / (C/D)

Relative risk = $(A/M_1) / (C/M_2)$

Note: In a case-control study, the total exposure is unknown because subjects were grouped according to disease of interest; therefore, M_1 and M_2 are unknown. Odds ratios are used as an estimation of risk.

Types of Effects

- By convention, assuming the intervention is the numerator and the comparator (which is often placebo) is the denominator:
  - ORs/RRs/HRs < 1 → protective/beneficial effect
    - $OR(stroke)_{ASA}$ = (221/19713) / (266/19676) = 0.829
    - $RR(stroke)_{ASA}$ = (221/19934) / (266/19942) = 0.831
  - ORs/RRs/HRs = 1 → NO effect
  - Ors/RRs/HRs > 1 → harmful/adverse effect

Ors/RRs/HRs and CIs

<u>Use of confidence intervals</u>

- Researchers often report a confidence interval around the relative risk rather than a p value.
- Finding is considered statistically significant based on <u>whether or not the CI contains 1</u>.
  - $RR(CVevent)_{ASA}$ = 0.91, 95% CI = 0.80 to 1.03
  - $RR(stroke)_{ASA}$ = 0.83, 95% CI = 0.69 to 0.99
  - $RR(hemstroke)_{ASA}$ = 1.24, 95% CI = 0.82 to 1.87
- If the CI contains 1, the relative risk is <u>not statistically significant</u>!

What is RRR?

- Relative risk reduction (RRR)
  - 1 – RR
- RRR (stroke)$_{ASA}$
  - 1 – RR = 1 – 0.83 = 0.17 or 17%

What is RRI?

- Relative risk increase (RRI)
  - |(1 – RR)|
  - calculated similarly to RRR
  - Again, perspective is important
    - Is the event beneficial or harmful?
    - Which is the intervention? Which is the comparator?

Please complete questions #2e – 2f and #3e – 3f on worksheet.

Relative vs. Absolute Risk Reduction/Increase

- RRRs and RRIs will almost always be larger than their corresponding ARRs and ARIs.
- Watch for these in drug ads!
  - RRRs may be used for efficacy outcomes
  - ARIs may be used for safety outcomes
  - Why is this?

Promotional Material: Risk

- Relative risk (RR) vs. absolute risk (AR) vs. number needed to treat (NNT)
- RR will always be higher than AR
- CURE study example
  - 20% RRR CV events (2.1% ARR)
  - 1% ARI in serious bleeds (38% RRI)

The CURE Trial Investigators. Effects of clopidogrel in addition to aspirin in patients with acute coronary syndromes without ST-segment elevation. N Engl J Med 2001; 345:494 – 502.

Promotional Material: Risk

For every 100 people treated with clopidogrel and ASA instead of clopidogrel alone, 2 will be spared a cardiovascular event and 1 will experience a serious bleed.

NNT = 1 / ARR
NNH = 1 / ARI

The CURE Trial Investigators. Effects of clopidogrel in addition to aspirin in patients with acute coronary syndromes without ST-segment elevation. N Engl J Med 2001; 345:494 – 502.

Important Point

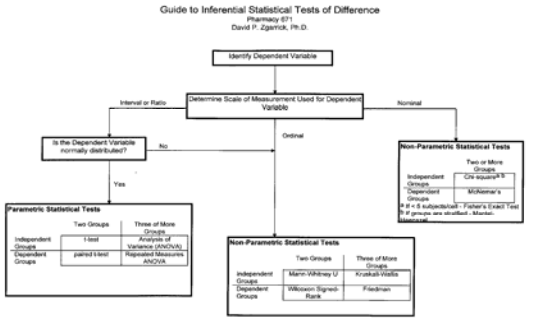
- RRs do not tell you anything about the overall incidence of the disease/outcome.
- For example:
  - Treatment X reduced mortality from 40% (with placebo) to 20% (with tx).
  - Treatment Y reduced mortality from 4% (with placebo) to 2% (with tx).
  - Relative risk of death with tx is 0.5 for both treatments.
  - Treatment X's NNT is 1/ARR = 1/0.20 = **5**
  - Treatment Y's NNT is 1/ARR = 1/0.02 = **50**

STATISTICAL TESTS

Statistical Analysis

- Statistical tests to be performed should always be determined before (*a priori)* the study takes place and should be described in the Methods section
- Your job
  - If reading a trial: Determine if the proper tests were performed. Interpret and evaluate the results.
  - If performing a study: Select and perform the appropriate test. Interpret and evaluate the results.

Statistics Flow Chart



Used with permission from Dr. David Zgarrick

One-tailed vs. Two-tailed Tests

One-tailed Test

- Research hypothesis states that one group will be higher/lower than another (directional)

Two-tailed Test

- Research hypothesis states that one group will be different than another (non-directional)
- Two-tailed tests are more common
  - Easier to reject null hypothesis

Two-tailed Tests

- Researchers often "cheat" when interpreting two-tailed tests
  - Assign a direction to the difference in the population based on the direction of the difference in the sample.
- This can result in "Type III" error
  - There is a statistically significant difference, but the direction of the difference stated by the researchers based on their sample data is incorrect.

Types of Groups

Independent Groups

- The subjects being compared in each group are different - mutually exclusive
- Also known as between-group comparisons
- Examples
  - RCT (parallel designs) – Drug A vs. Drug B, Drug vs. Placebo
  - Males vs. Females
  - P-1 vs. P-2 vs. P-3 vs. P-4

Types of Groups

Dependent Groups

- The subjects being compared in each group are the same
- Also known as within-group comparisons
- Examples
  - Pretest - Posttest
  - Before - After
  - Crossover designs

Parametric vs. Non-parametric

- Parametric
  - More "powerful" than non-parametric tests if used appropriately
  - Appropriate for NORMALLY distributed data
  - Rule of thumb: if sample size ≥ 30 pieces of continuous data, it's likely to be normally distributed
  - Appropriate for continuous data
- Non-parametric
  - Less "powerful" than parametric tests
  - Appropriate for non-normally distributed data
  - Appropriate for nominal and ordinal data

REMEMBER PARABOLA!

A researcher designs a study in which 3 HMGCoA Reductase Inhibitors are compared in 3 separate groups of 100 subjects (total 300). The primary outcome measure is LDL at 8 weeks, and the null hypothesis is that there is no difference in the treatments with respect to LDL.

Which statistical test should be used to analyze the data?

A – t-test
B – paired t-test
C – ANOVA
D – Mann Whitney U
E – Chi-square

Putting It All Together

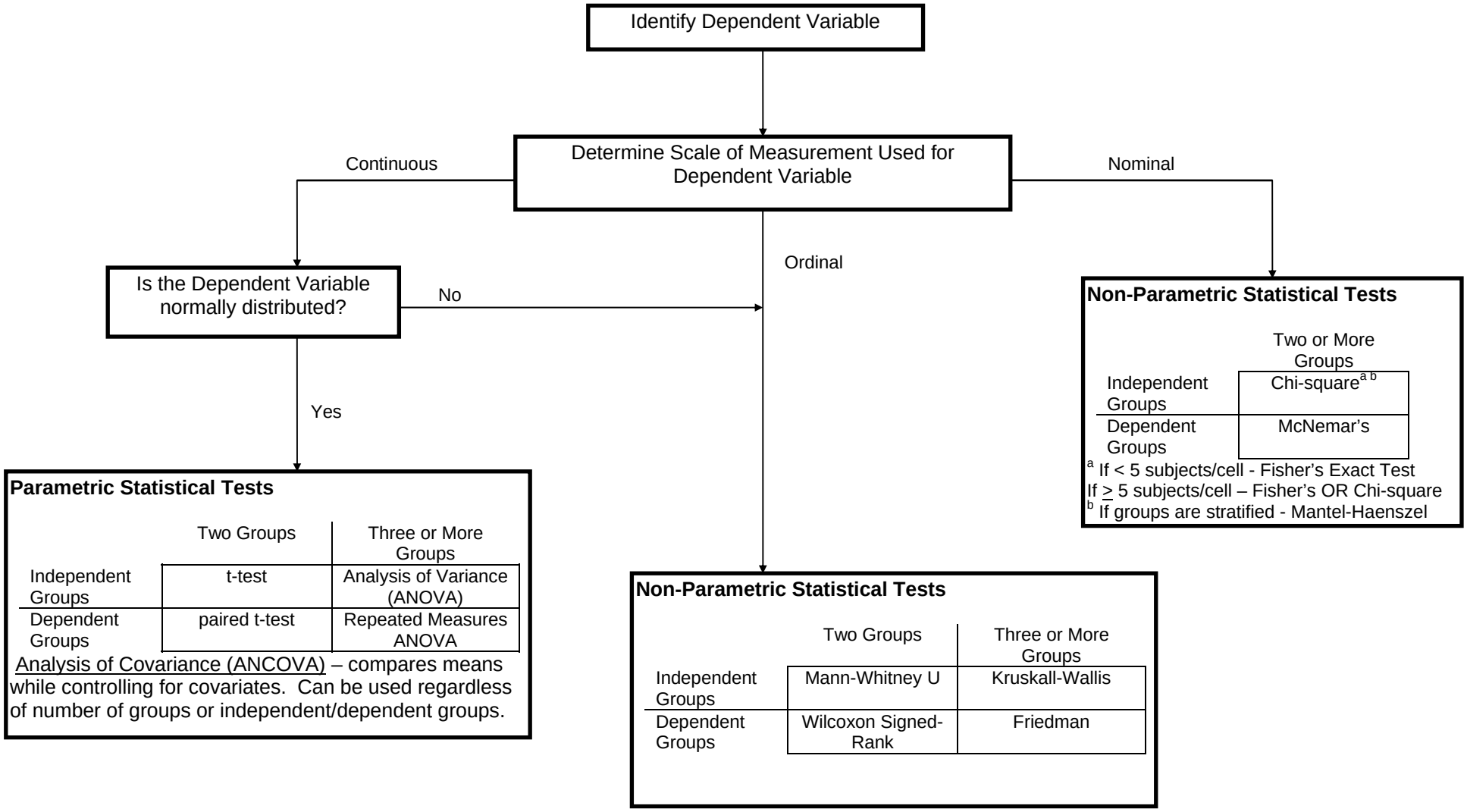
- How can all this information used?
  - Better conceptualize risk and benefit in clinical trials and pharmaceutical advertisements
  - Better apply the results of studies to patient care
    - Considering the worksheet:
      - What if your patient was at very high risk for stroke or MI?
      - What if your patient was at very high risk for bleeding?
      - How would this affect your choice whether clopidogrel should be added to ASA?

Statistics in Clinical Practice

Jacob P. Gettig, PharmD, MPH, BCPS
Assistant Dean for Postgraduate Education
Associate Professor of Pharmacy Practice
Midwestern University Chicago College of Pharmacy

Guide to Inferential Statistical Tests of Difference

David P. Zgarrick, Ph.D.



Parametric Statistical Tests

	Two Groups	Three or More Groups
Independent Groups	t-test	Analysis of Variance (ANOVA)
Dependent Groups	paired t-test	Repeated Measures ANOVA

Analysis of Covariance (ANCOVA) – compares means while controlling for covariates. Can be used regardless of number of groups or independent/dependent groups.

Non-Parametric Statistical Tests

	Two Groups	Three or More Groups
Independent Groups	Mann-Whitney U	Kruskall-Wallis
Dependent Groups	Wilcoxon Signed-Rank	Friedman

Non-Parametric Statistical Tests

	Two or More Groups
Independent Groups	Chi-square[a b]
Dependent Groups	McNemar's

[a] If < 5 subjects/cell - Fisher's Exact Test
If ≥ 5 subjects/cell – Fisher's OR Chi-square
[b] If groups are stratified - Mantel-Haenszel

Important
Independent Groups - Different subjects in each group (mutually exclusive)
Dependent Groups - Same subjects in each group (paired/matched/repeated measures)

Distributed with permission from the author, Dr. David Zgarrick.

Statistics in Clinical Practice
ICHP Annual Meeting 2009
Jacob Gettig, PharmD, MPH, BCPS

In groups of 2 to 4, please answer the following questions.

Table 4. Composite and Individual Primary and Secondary End Points.

End Point	Clopidogrel plus Aspirin (N=7802)	Placebo plus Aspirin (N=7801)	Relative Risk (95% CI)*	P Value
	no. (%)			
Efficacy end points				
Primary efficacy end point	534 (6.8)	573 (7.3)	0.93 (0.83–1.05)	0.22
Death from any cause	371 (4.8)	374 (4.8)	0.99 (0.86–1.14)	0.90
Death from cardiovascular causes	238 (3.1)	229 (2.9)	1.04 (0.87–1.25)	0.68
Myocardial infarction (nonfatal)	146 (1.9)	155 (2.0)	0.94 (0.75–1.18)	0.59
Ischemic stroke (nonfatal)	132 (1.7)	163 (2.1)	0.81 (0.64–1.02)	0.07
Stroke (nonfatal)	150 (1.9)	189 (2.4)	0.79 (0.64–0.98)	0.03
Secondary efficacy end point†	1301 (16.7)	1395 (17.9)	0.92 (0.86–0.995)	0.04
Hospitalization for unstable angina, transient ischemic attack, or revascularization	866 (11.1)	957 (12.3)	0.90 (0.82–0.98)	0.02
Safety end points				
Severe bleeding	130 (1.7)	104 (1.3)	1.25 (0.97–1.61)	0.09
Fatal bleeding	26 (0.3)	17 (0.2)	1.53 (0.83–2.82)	0.17
Primary intracranial hemorrhage	26 (0.3)	27 (0.3)	0.96 (0.56–1.65)	0.89
Moderate bleeding	164 (2.1)	101 (1.3)	1.62 (1.27–2.08)	<0.001

* CI denotes confidence interval.
† The secondary efficacy end point was the first occurrence of myocardial infarction, stroke, death from cardiovascular causes, or hospitalization for unstable angina, a transient ischemic attack, or a revascularization procedure (coronary, cerebral, or peripheral).

Adapted from: Bhatt DL et al. *N Engl J Med*. 2006; 354(16): 1706-17.

1. Assuming alpha was set at 0.05, which efficacy and safety end points were statistically significant? How do you know?

2. For the secondary efficacy endpoint, identify/calculate the following:

    a. $AR_{clopidogrel+ASA}$

    b. $AR_{placebo+ASA}$

    c. ARR

    d. NNT

    e. RRR

    f. 95% CI for the RRR

3. For the safety endpoint of moderate bleeding, identify/calculate the following:

    a. $AR_{clopidogrel+ASA}$

    b. $AR_{placebo+ASA}$

    c. ARI

    d. NNH

    e. RRI

    f. 95% CI for the RRI

4. Make a statement regarding the risk/benefit trade-off using the NNT and NNH for the above efficacy and safety endpoints.