Why discuss statistics ?

- To understand clinical research studies in journals.
- To design clinical research studies.
- To analyze clinical research studies.
- To be better able to explain results of clinical research studies to our patients.
- To answer a few questions on tests.

Types of Clinical Research Studies

- Cohort: patients have some condition/something in common
- Case-Control: cases have some condition; controls don't
  - Often aspect of cohort study, where controls are 'matched' with cases in cohort for age, gender, and sometimes other variables such as date of admission or date of encounter
- Randomized, Placebo-controlled or Sham Treatment-Controlled : all patients have the condition; treatment determined by chance
  - May be unblinded, single blinded or double blinded
- Randomized, Active-treatment controlled trial: all patients have the condition; treatment determined by chance
  - May be unblinded, single blinded or double blinded
  - Often a phase 3 trial; may be non-inferiority study design
- Meta analysis: pooling multiple separate studies of some condition, although definitions of condition and outcome of interest may vary from study to study; meta has many variations

Two Types of Variables in Clinical Research Studies

CONTINUOUS

- AGE
- Blood Pressure
- Serum AST
- Serum CRP
- Fasting blood glucose
- HEIGHT/WEIGHT/BMI

CATEGORICAL (2 or more)

- GENDER
- RACE
- DIABETIC?
- PREGNANT?
- CURE?
- OLD vs. YOUNG

Different statistical methods are used with continuous vs. categorical variables. Continuous variables can be normally distributed (bell shaped) or skewed. Different tests are used for normal versus skewed data.

Basic Statistical Terms Describing Data

- Data set: Test scores on a 20-question exam in 21 students:
  {**2**,13,4,**20**,18,6,6,11,**9**,12,5,4,8,18,10,11,4,**20**,16,7,5}
- Range: the **extreme** values (min and max) = 2 to 20
- Median: the **middle** value, dividing the population into 2 subgroups; 11^{th} value =9
  - Quartiles: divides **all** values into 4 groups (1^{st},2^{nd},3^{rd},4^{th})
    - Used to establish the Interquartile range (values spanning 2^{nd} and 3^{rd} Q)
  - Tertiles (3 groups), Quintiles (5 groups), Percentiles (100 groups)
- Mean: average value (uses **all** values) = $\sum$21 scores/21= 10
  - Mean is usually cited along with the standard deviation of the mean

Standard deviation of the mean

- Measures the average of the differences from the mean value among the values in the data set:

$$SD = \sqrt{(\Sigma(\text{differences from the mean}^2)/n-1)}$$

- A statistic to deal with values higher than the mean (+ difference,+d) and values lower than the mean (- difference,-d). $(-d)^2=(+d)^2=d^2$. And $\sqrt{d^2}=x$
- Applies to continuous data that are more or less normally distributed (bell shaped curve)

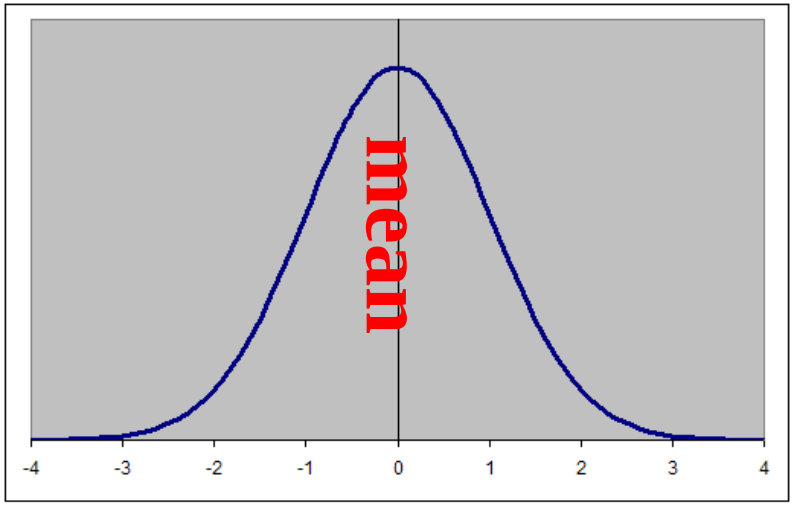
Example: standard deviation calculation

Values	Difference from Mean (d)	d^2
12	+2	4
10	0	0
5	-5	25
15	+5	25
8	-2	4
$\sum$=50/5=10		$\sum=d^2/(n-1)$= 58/4=14.5
Mean,10	Differences can be + or -	$\sqrt{14.5}$ = 3.8 = SD

Mean ± SD = 10 ± 3.8

The normal (bell-shaped) distribution

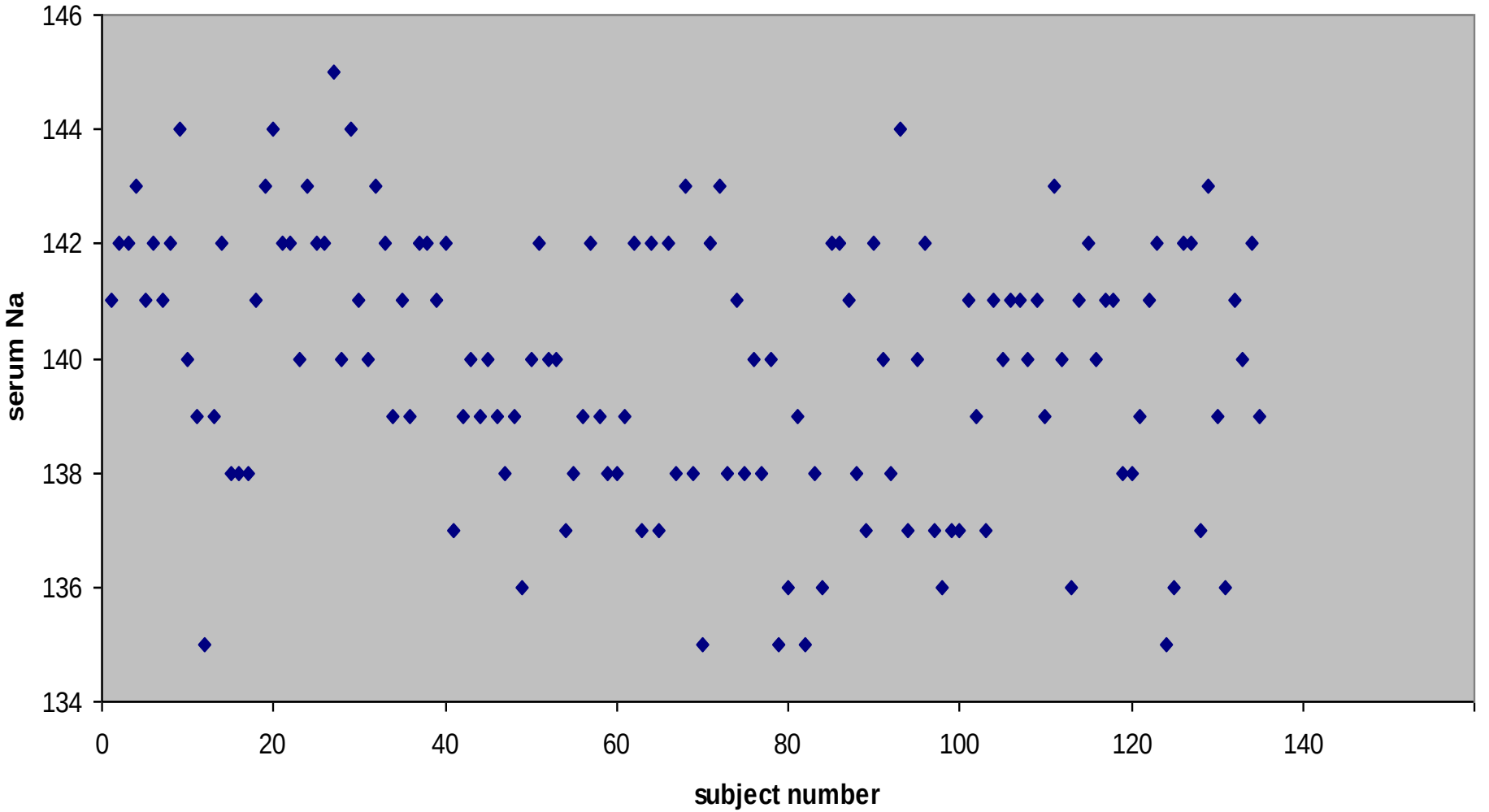
n



Standard deviations (SD) from the mean.
(95% of values are within 1.96 SD of mean.)

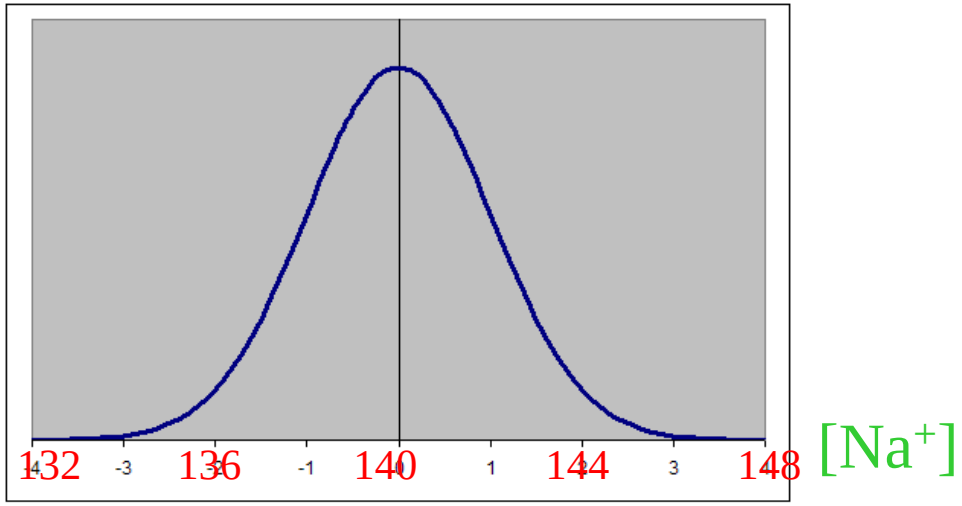
Serum [Na^+] in 135 normal people

(Soni N and Feldman M)



Mean, 140; median 140; range, 135-145 mM; standard deviation 2.0 mM

The normal serum [Na^+] distribution



In the case of serum Na, mean is 140 and SD is 2.0 and thus ~95% of values are between 140 minus 4 and 140 plus 4, or between 136 -144.

Some important statistical concepts

- **Confidence intervals (CIs, usually reported as 95% CIs)**
- **Absolute risk reduction and relative risk reduction**
- **Number needed to treat/ number needed to harm**
- **Type 1 and Type 2 errors**
- **Estimating sample size when designing a study**

- t tests
- 2-by-2 tables (Chi square, Fisher exact, others)
- Odds ratios or hazard ratios
- Pre- and post-test probabilities and likelihood ratios
- Non-inferiority study designs
- Survival models and Cox proportional ratios

Ann Int Med 2009: 150: JC6-16

95% Confidence intervals (CIs)

Concept: How **confident** can we be about the size of a *measured* difference and whether the measured difference is statistically significant (i.e., it would be expected to occur less than 5% of the time by chance alone)

Example: 95% CI calculation

H. pylori eradication/NSAID study* of 100 patients, with a *categorical* outcome: ulcer vs. no ulcer.

- Group 1: 5 of 51 (10%, or .10) pts. randomized to receive antibiotics for *H. pylori* got ulcers when given an NSAID. Let's call .10 p_1.
- Group 2: 15 of 49 (31%, or .31) pts. randomized not to receive antibiotics got ulcers when given an NSAID. Let's call .31 p_2.

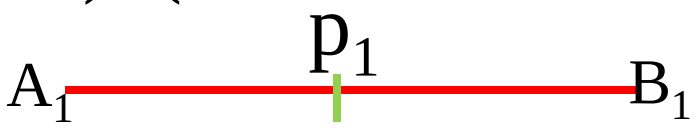
*Lancet 2002; 359:9-13

95% CIs

The proportions, $\mathbf{p_1}$ and $\mathbf{p_2}$, of patients who got ulcers in each group are *estimates* of the true proportions who would get ulcers.

From these *estimates*, we can be 95% confident that the actual ulcer rates range from A_1 to B_1 and A_2 to B_2, with p_1 and p_2 in the center of the interval from A to B.

A_1 and B_1 represent the *95% confidence intervals* for p_1 (see red line). (A similar relationship for p_2.)



95% Confidence interval (CI) formula for a proportion, p:

$$95\%\ CI = p \pm 1.96\sqrt{[(p)(1-p)/n]}$$

(p)(1-p) can range from close to 0.00 to 025. The closer p is to 0.5, the larger (p)(1-p). The larger the n, the smaller (more precise) the CI.

H. pylori/NSAID study

5 of 51 (p_1=10%, or .10) of the antibiotic group got ulcers when exposed to NSAID

- **95% CI =.10 ± 1.96$\sqrt{(.1)(.9)/51}$=.10 .08=[.02, .18]→ [2%,18%]**

15 of 49 (p_2=31%, or .31) of the placebo- group got ulcers when exposed to NSAID

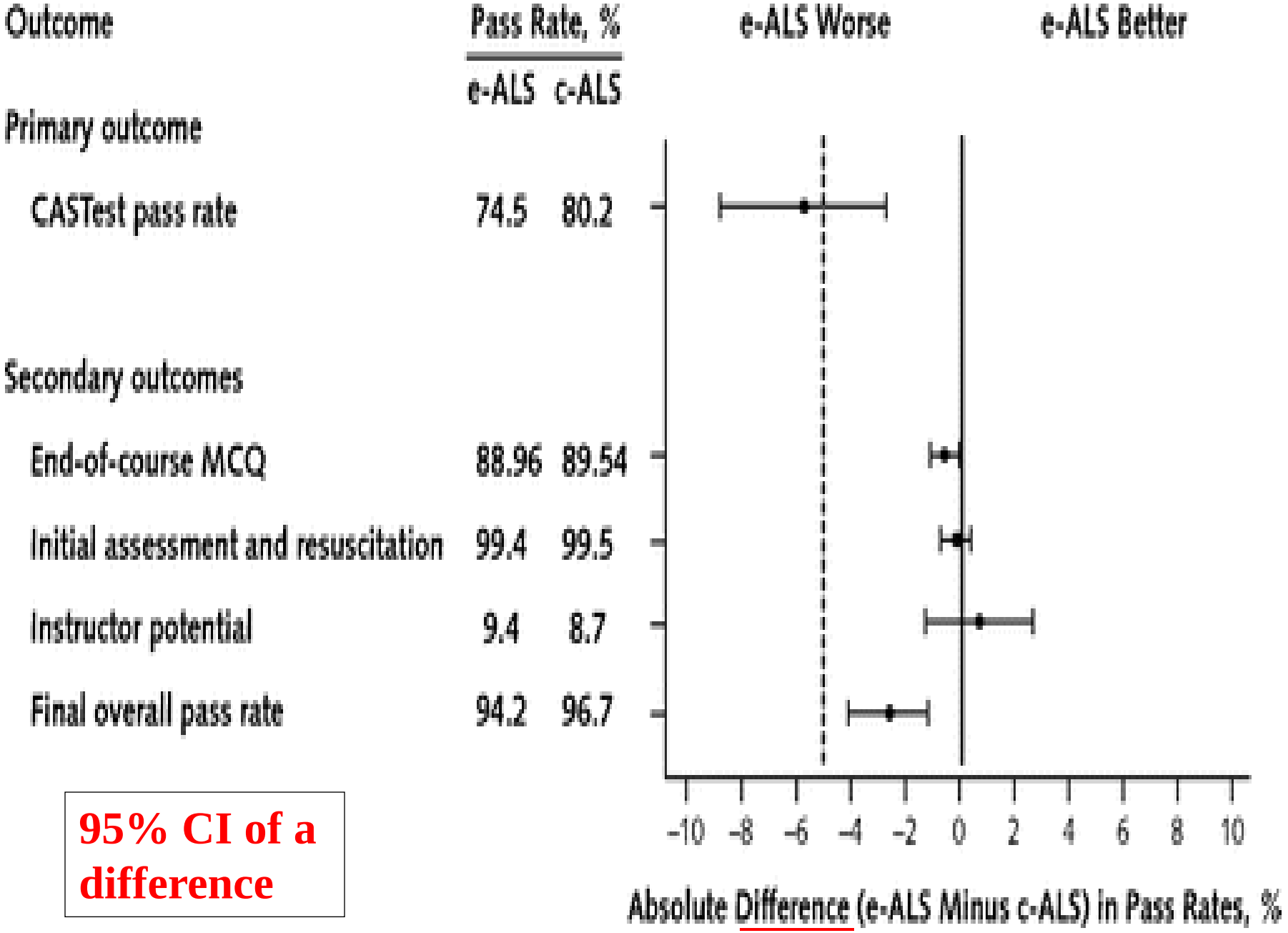
- **95%CI =.31±1.96$\sqrt{(.31)(.69)/49}$ =.31 .13=[.18,.44]→[18%, 44%]**

Fisher exact test (to be discussed in Statistics 102) to assess signifance in ulcer rate (categorical variable) : P = 0.0122

http://www.graphpad.com/quickcalcs/index.cfm

95% CI of a difference in systolic BP

- Two antihypertensive medications are compared in 400 patients with mild systolic hypertension, 198 of whom were randomly treated with HCTZ and 202 with a new diuretic. Mean systolic BP at the beginning of the study was 144 ±8 mmHg in both groups. By the end of 8 weeks, mean ± SD systolic BP was 138±8 mmHg in the HCTZ group and 135±9 mmHg in the new group.
- What is 95% CI of this 3 mmHg difference, and does it include (overlap with) zero? Use unpaired t test to determine significance of the BP change (continuous variable).
- http://www.graphpad.com/quickcalcs/index.cfm



95% CI of a difference

Absolute Risk Reduction (ARR) (and its 95% CI)

- Back to our *H. pylori* paper:
- The ARR with antibiotics, p_1-p_2, was .31-.10 or .21
- The 95% CI of this ARR =

  $(p_1-p_2) \pm 1.96 \sqrt{(p_1)(1-p_1)/n_1 + (p_2)(1-p_2)/n_2}$ = .21±.15 or [6%, 36%].
- The ARR with antibiotics is somewhere between 6% and 36%, with 95% confidence.
- This CI does not overlap zero and thus is unlikely due to chance.

Relative Risk Reduction (RRR)

- Relative Risk Reduction (RRR)=ARR/Risk with placebo (or control).
- In the H. pylori example, RRR= .21/.31 = .68=68%. What this means:
  - If we were to treat 1,000 pts. with NSAID→ 310 ulcers expected (31%)
  - If we were to treat 1,000 pts. with NSAID + Abs→ 100 ulcers expected (10%)
  - Thus, antibiotic therapy would have prevented 210 ulcers
  - And 210 ulcers prevented÷310 ulcers expected = 68% ulcer prevention rate = RRR.
  - Antibiotic use reduced the # of ulcers from 310 to 100, or to 32% of expected, a RRR of 68%.
- Note: Length of exposure to NSAID in this study in the 2 groups was identical.

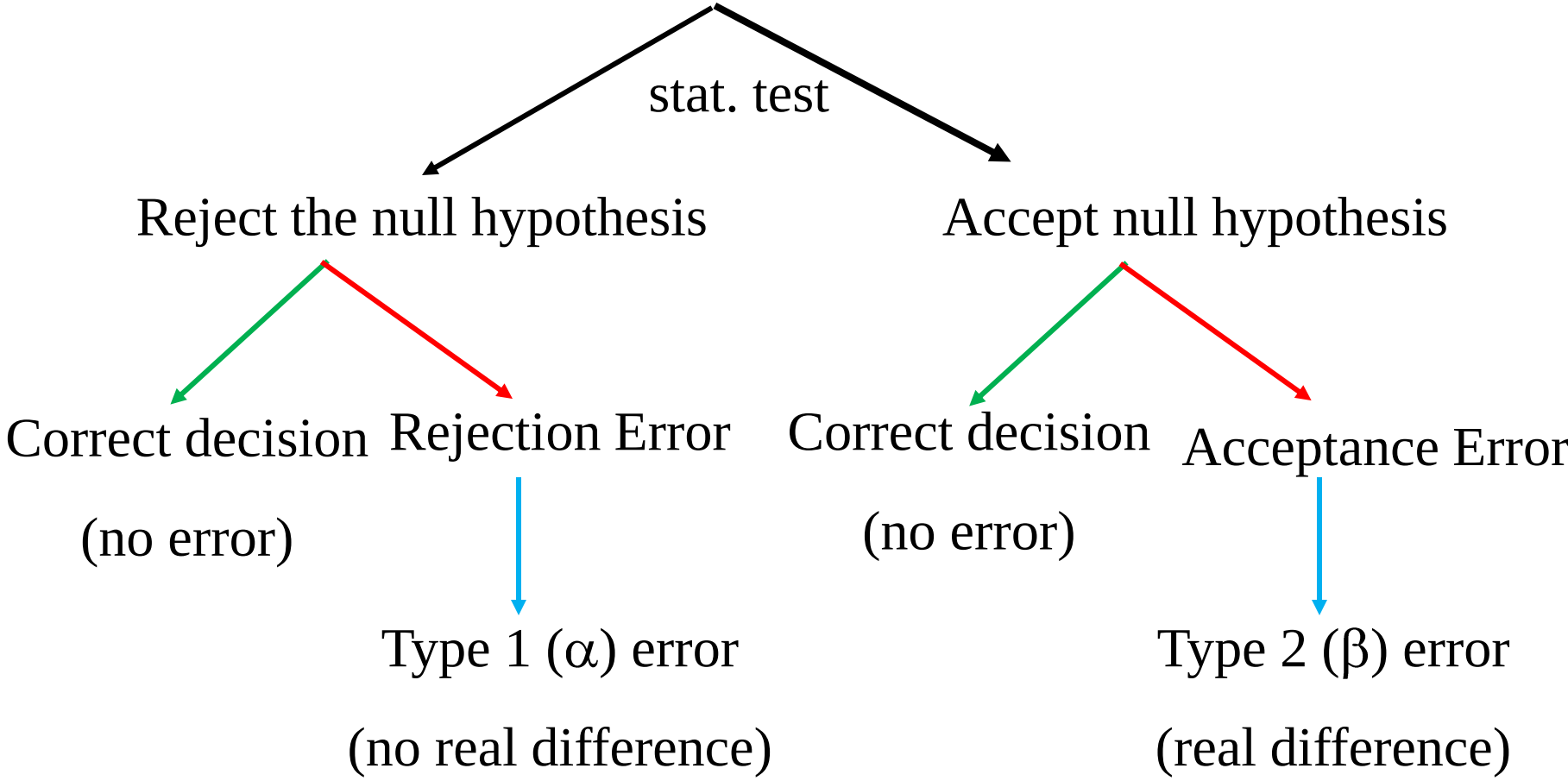
  If the 2 groups had not been followed for an identical time, often the case in trials, outcomes may be higher in the group followed longer and thus events would need to be expressed per unit of time (e.g., events per 100 patient-years).

The number needed to treat (NNT)

- NNT= 1/ARR
- Since the Absolute Risk Reduction (ARR) in our study was .31-.10= .21, then the NNT= 1/.21≅ 5.
- The number needed to harm (NNH) uses the same concept as NNT, except that the intervention caused harm rather than good.
- It is easy to determine that the 95% CI of NNT is 3 to 18:
  - http://www.graphpad.com/quickcalcs/index.cfm

Type 1(α) and Type 2 (β) Errors

Null Hypothesis: there is no differences between two treatments



Selecting the **size** of α and β errors

- The type 1 error, or α (also called the P value) is conventionally set below 0.05 (5%)
  - i.e, chance of a type 1 error <u>if</u> the null hypothesis is rejected is < 5%
  - Can state "p<0.05" or give exact p value (e.g., p=0.001, or p=0.049)
- The type 2 error, or β, is often set at 2 to 4 times α , or 0.10-0.20 (10%-20%)
  - i.e., chance of making a type 2 error <u>if</u> the null hypothesis is accepted is 10-20%
  - <u>POWER</u> to detect a real difference (and thus reject the null hypothesis) = 1- β
  - smaller β (e.g., 0.1), more power (.9)
  - larger β (e.g., 0.2), less power (.8)
- If a study is highly powered and the null hypothesis is accepted, the chance of there being a true difference is quite small.
- If the study is under-powered and the null hypothesis is accepted, there can be little confidence that a true difference has been excluded.

Use of α and β in sample size planning/study design

A new antibiotic is developed for *C. difficile*.

How many patients would be needed to be included in a phase 3 trial to be able to show that this new drug is superior to metronidazole?

To answer this question, we need to know:

1. What is the expected success rate for metronidazole? [p_1]

2. What would be a clinically important and expected improvement in success rate (based on phase 1 and 2 studies) with the new drug? [p_2]

3. What should be the α (type 1 error) and the β (type 2 error) for the study? (Recall: Power = 1- β.)

Sample size estimation, cont'd

- p1 = 0.75 (metronidazole, based on literature/prior studies)

  p2 = 0.90 (New Rx, based on small, initial phase 1 and 2 trials)

  α = 0.05 (<1 in 20)

  β = 0.10 (1 in 10). Power = 0.90 (9 in 10)

- According to standard tables (Fleiss), we would need 158 patients per group, or 316 patients in total for this α and power.

- If ~10% drop out rate is expected, then 158+16=174 per group, or 348 patients in total would need be randomized.

  (This sample size may necessitate a multi-center study to enroll sufficient patients during the proposed time frame.)

- Analyze data by intent-to-treat and by evaluable patients. Primary outcome: cure, which is categorical variable. Probably would use Fisher exact test as 2 by 2 test.