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

- **2-by-2 tables (Chi square, Fisher exact, others)**
- **Odds ratios or Hazard ratios**

- **Sensitivity, Specificity and Receiver Operator Curves**
- **Likelihood ratios and Positive/Negative predictive Values**

- **Tests to Assess Statistical Significance (p values)**

- **Non-inferiority study designs**

Ann Int Med 2009: 150: JC6-16

2x2 Contingency Tables:
Chi Square/Fisher Exact /etc.

(used for categorical outcomes to calculate P values and odds ratios)

- A new treatment for Crohn's disease is compared to a standard treatment in 245 patients.
- **120** patients are randomized to the new treatment and **125** to the standard treatment, each for eight weeks.
- **90/120** given the new treatment group go into remission (75%) and 30/125 (25%) do not.
- **75/125** given the standard treatment go into remission (60%) and 50/125 (40%) do not.
- Remission (categorical variable) pre-defined as CDAI.
- Was there a significant improvement in outcome, or could this outcome have been due to chance? ……Let's vote!

Step 1: create standard 2x2 table

	REMIT	NO REMIT
New Rx (a+b)	a	b
Standard Rx (c+d)	c	d

Enter the data from our study

	REMIT	NO REMIT
New Rx (n=120)	90(a)	30(b)
Standard Rx (n=125)	75(c)	50(d)

Chi square (χ^2) test

$$\chi^2 = \frac{n\,(|ad-bc| - n/2)^2}{(a+b)(c+d)(a+c)(b+d)}$$

$\chi^2 = 6.264$ (p=0.012)

Fisher exact test: p=0.014

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

Odds ratio (OR) of a remission

New Rx	90 (a)	30 (b)
Standard Rx	75 (c)	50 (d)

a+b+c+d=n=total patients in study
a/b = odds of remission with New Rx; 3:1
c/d= odds of remission with Standard Rx; 1.5:1
a/b÷c/d= odds ratio of New compared to Standard Rx=ad/bc
Odds ratio = 4,500/ 2,250= **2.0;** or 3:1÷1.5:1=**2.**

This odds ratio of 2.0 might have occurred by chance alone.*
95% CI of the odds ratio or hazard ratio:

* We know it did not occur by chance alone due to chi square/Fisher test results.

95% CI of an odds ratio

THE BASICS:

- **$\log_{10}x$=the power by which you must raise 10 to obtain x.**
- **$\log_{10}100$=2 because 10^2=100; $\log_{10}10$=1 because 10^1=10 and $\log_{10}1$=0.**
- **e≅ 2.71828182846**
- **$\log_e x$ or ln$x$= the power by which you must raise the number e in order to obtain x.**
- **ln2.71828182846=1 and ln 1=0.**
- **Thus, if ad/bc =1, then ln ad/bc=0**
- **If ad/bc>1, ln ac/b>0 (i.e., is a positive number, such as 0.13 or 6.98)**
- **If ad/bc<1, ln ad/bc<0)i.e., is a negative number, such as -0.47 or -3.01)**

Calculating 95% CI of the odds ratio (OR)

- Step 1: Calculate the ln of the 95% CI:

ln 95% CI = ln ad/bc $\pm$ 1.96$\sqrt{}$ 1/a+1/b+1/c+1/d

Colitis study: ln 95% CI = ln 2.0 $\pm$ 1.96$\sqrt{}$ 1/90+1/30+1/75+1/50

Since ln 2.00= 0.693

Thus, ln 95% CI= 0.693 $\pm$ 0.508 = (+0.185, +1.201).

- Step2: From the ln of the 95% CI, determine the 95% CI.

To find the actual 95% CI for the OR, we must find the antiln of +0.185 and of +1.201.

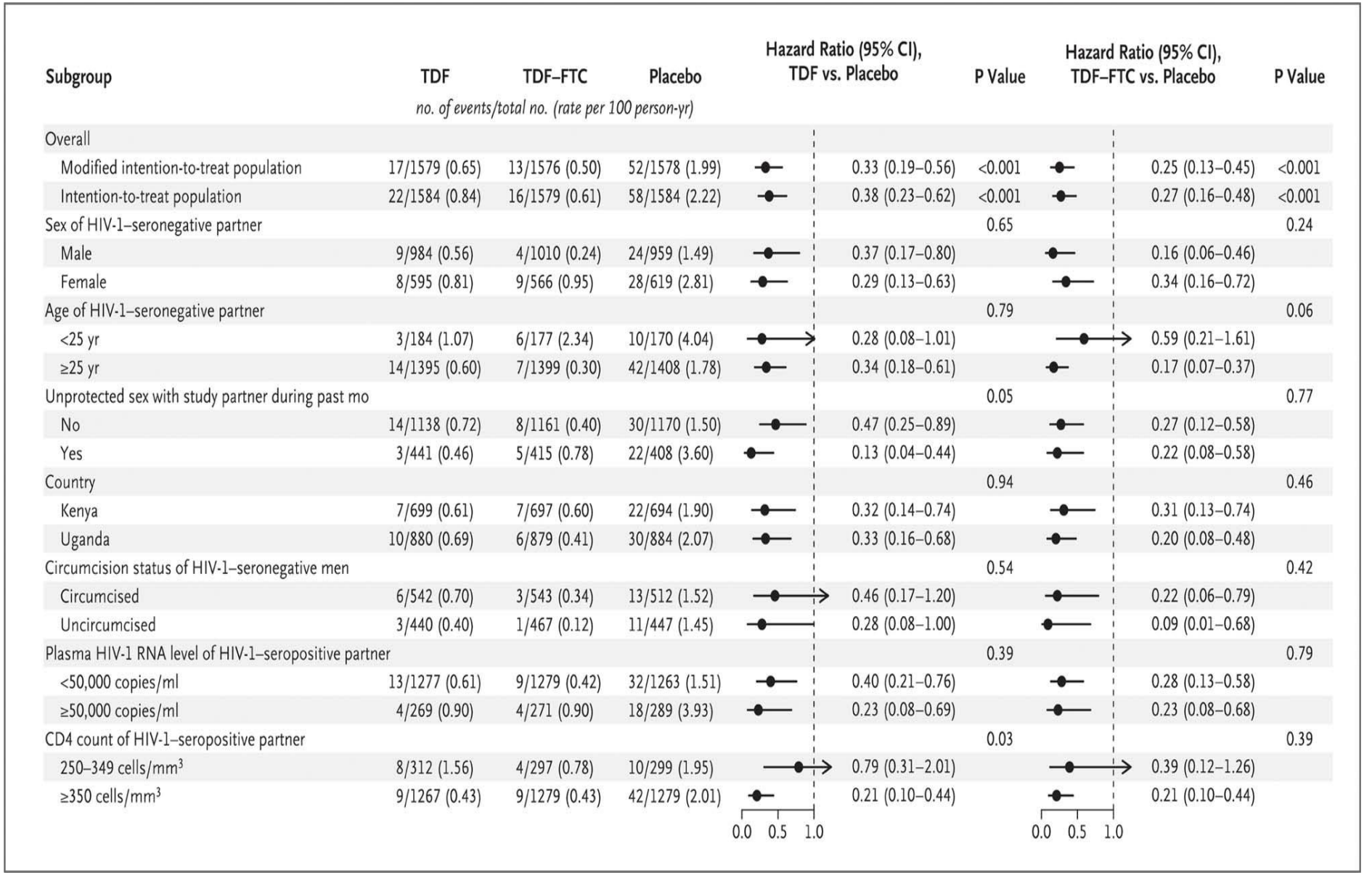
Antiln x is the number that results when you raise e to the x power.

antiln +0.185 = $e^{.185}$ =**1.20**

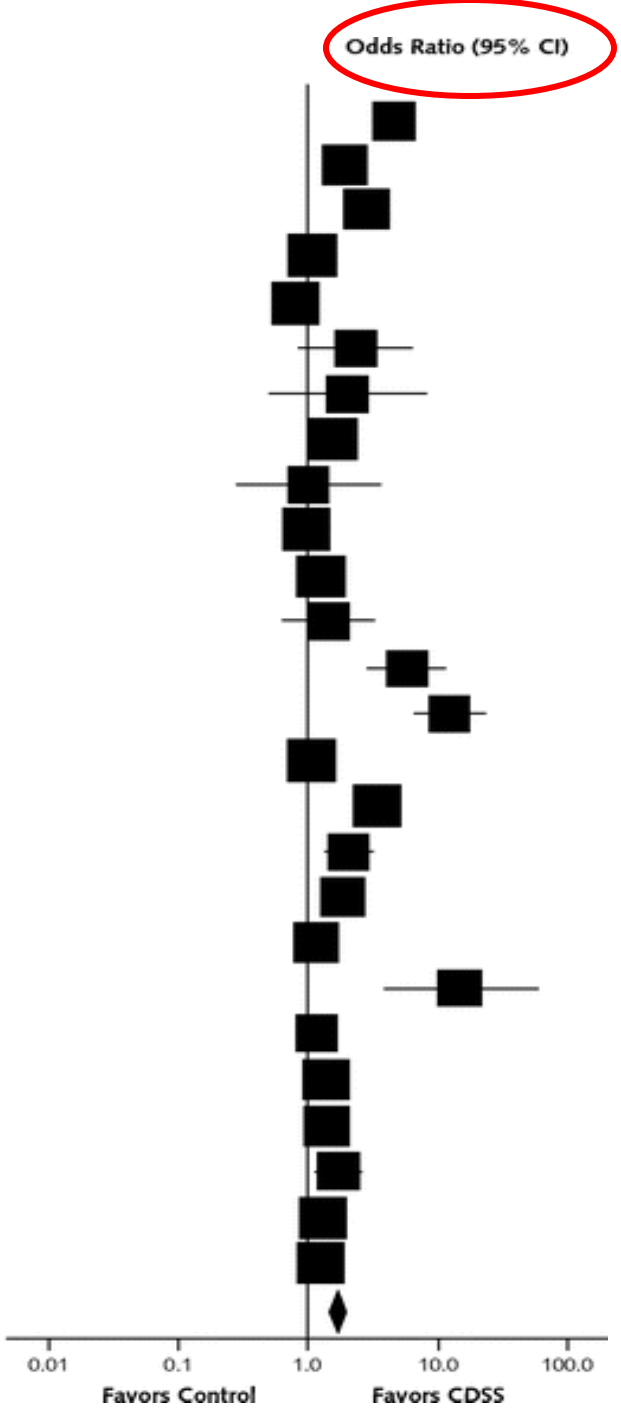
antiln +1.201 = $e^{1.201}$ =**3.32**.

$\therefore$ 95% CI of the OR =**1.20, 3.32.**

Thus, the odds ratio for a remission with the new treatment is **2.00 (95% CI, 1.20-3.32)**. As this odds ratio does not cross 1.00, the difference is unlikely due to chance and is significant at the 0.05 level.


Subgroup
TDF
TDF–FTC
Placebo
no. of events/total no. (rate per 100 person-yr)
Hazard Ratio (95% CI), TDF vs. Placebo
P Value
Hazard Ratio (95% CI), TDF–FTC vs. Placebo
P Value
Overall
Modified intention-to-treat population 17/1579 (0.65) 13/1576 (0.50) 52/1578 (1.99) 0.33 (0.19–0.56) <0.001 0.25 (0.13–0.45) <0.001
Intention-to-treat population 22/1584 (0.84) 16/1579 (0.61) 58/1584 (2.22) 0.38 (0.23–0.62) <0.001 0.27 (0.16–0.48) <0.001
Sex of HIV-1–seronegative partner 0.65 0.24
Male 9/984 (0.56) 4/1010 (0.24) 24/959 (1.49) 0.37 (0.17–0.80) 0.16 (0.06–0.46)
Female 8/595 (0.81) 9/566 (0.95) 28/619 (2.81) 0.29 (0.13–0.63) 0.34 (0.16–0.72)
Age of HIV-1–seronegative partner 0.79 0.06
<25 yr 3/184 (1.07) 6/177 (2.34) 10/170 (4.04) 0.28 (0.08–1.01) 0.59 (0.21–1.61)
≥25 yr 14/1395 (0.60) 7/1399 (0.30) 42/1408 (1.78) 0.34 (0.18–0.61) 0.17 (0.07–0.37)
Unprotected sex with study partner during past mo 0.05 0.77
No 14/1138 (0.72) 8/1161 (0.40) 30/1170 (1.50) 0.47 (0.25–0.89) 0.27 (0.12–0.58)
Yes 3/441 (0.46) 5/415 (0.78) 22/408 (3.60) 0.13 (0.04–0.44) 0.22 (0.08–0.58)
Country 0.94 0.46
Kenya 7/699 (0.61) 7/697 (0.60) 22/694 (1.90) 0.32 (0.14–0.74) 0.31 (0.13–0.74)
Uganda 10/880 (0.69) 6/879 (0.41) 30/884 (2.07) 0.33 (0.16–0.68) 0.20 (0.08–0.48)
Circumcision status of HIV-1–seronegative men 0.54 0.42
Circumcised 6/542 (0.70) 3/543 (0.34) 13/512 (1.52) 0.46 (0.17–1.20) 0.22 (0.06–0.79)
Uncircumcised 3/440 (0.40) 1/467 (0.12) 11/447 (1.45) 0.28 (0.08–1.00) 0.09 (0.01–0.68)
Plasma HIV-1 RNA level of HIV-1–seropositive partner 0.39 0.79
<50,000 copies/ml 13/1277 (0.61) 9/1279 (0.42) 32/1263 (1.51) 0.40 (0.21–0.76) 0.28 (0.13–0.58)
≥50,000 copies/ml 4/269 (0.90) 4/271 (0.90) 18/289 (3.93) 0.23 (0.08–0.69) 0.23 (0.08–0.68)
CD4 count of HIV-1–seropositive partner 0.03 0.39
250–349 cells/mm3 8/312 (1.56) 4/297 (0.78) 10/299 (1.95) 0.79 (0.31–2.01) 0.39 (0.12–1.26)
≥350 cells/mm3 9/1267 (0.43) 9/1279 (0.43) 42/1279 (2.01) 0.21 (0.10–0.44) 0.21 (0.10–0.44)
0.0 0.5 1.0
0.0 0.5 1.0

Author, Year (Reference)	Odds Ratio (95% CI)
McDonald, 1976 (94)	4.640 (3.197–6.734)
McDowell et al, 1989 (95)	1.930 (1.395–2.670)
Bates et al, 1999 (83)	2.870 (2.180–3.779)
Flottorp et al, 2002a (88)	1.100 (1.004–1.205)
Flottorp et al, 2002b (88)	0.810 (0.729–0.899)
Greiver et al, 2005a (89)	2.370 (0.833–6.744)
Greiver et al, 2005b (89)	2.040 (0.492–8.461)
Raebel et al, 2005 (99)	1.600 (1.439–1.779)
Tierney et al, 2005 (26)	1.020 (0.278–3.738)
Palen et al, 2006 (96)	0.980 (0.941–1.021)
Raebel et al, 2006 (100)	1.280 (1.179–1.389)
Wilson et al, 2006 (109)	1.460 (0.628–3.392)
Roukemia et al, 2008 (101)	5.860 (2.828–12.142)
Lee et al, 2009 (91)	12.540 (6.481–24.264)
Lo et al, 2009 (92)	1.070 (0.935–1.224)
Roy et al, 2009 (102)	3.450 (2.800–4.250)
Schriefer et al, 2009 (103)	2.070 (1.314–3.260)
Sundaram et al, 2009 (105)	1.880 (1.373–2.575)
Bell et al, 2010a (84)	1.160 (0.894–1.506)
Bell et al, 2010b (84)	15.290 (3.752–62.301)
Khan et al, 2010a (31)	1.170 (0.798–1.716)
Khan et al, 2010b (31)	1.390 (1.077–1.794)
Khan et al, 2010c (31)	1.400 (1.063–1.845)
Khan et al, 2010d (31)	1.740 (1.128–2.685)
Player et al, 2010 (98)	1.330 (1.132–1.563)
Walker et al, 2010 (108)	1.270 (1.111–1.452)
	1.716 (1.472–2.001)



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

- 2-by-2 tables (Chi square, Fisher exact, others)
- Odds ratios or hazard ratios

- **Sensitivity, Specificity and Receiver Operator Curves/AUCs**
- **Likelihood ratios and Positive/Negative predictive Values**

- **Tests to Assess Statistical Significance (p values)**

- **Non-inferiority study designs**

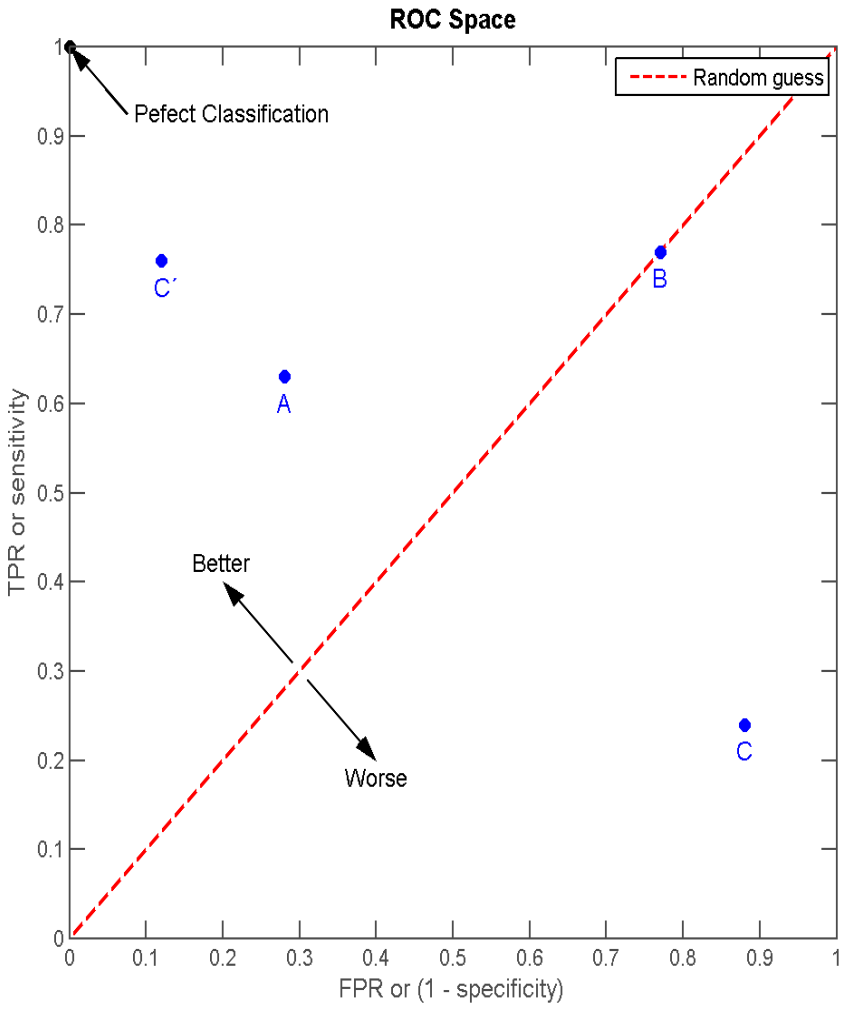
Ann Int Med 2009: 150: JC6-16

Sensitivity and Specificity

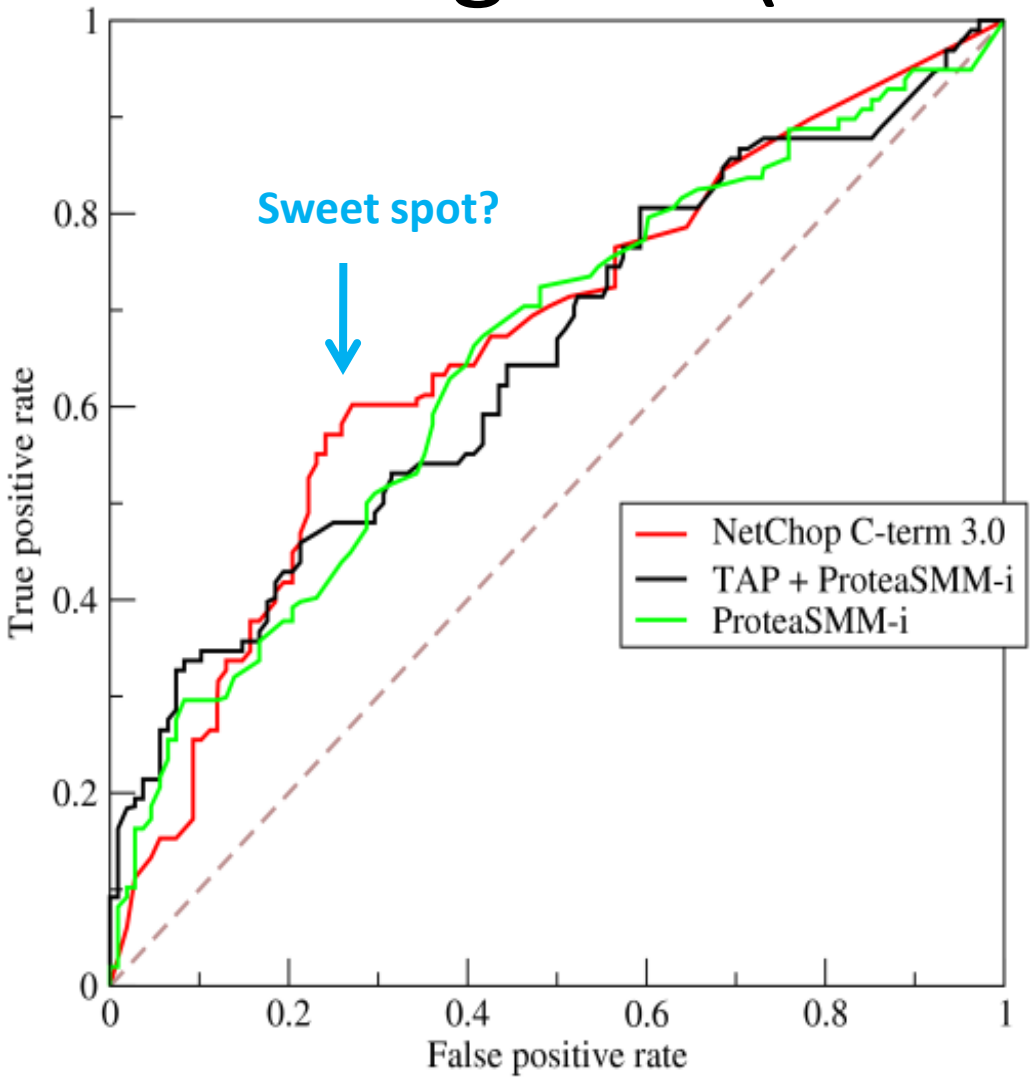
- Sensitivity: true positives (proportion of individuals with the disease who test +; ranges from 0 to 1, or from 0% to 100%)
- 1-Sensitivity: false negatives (proportion of individuals with the disease who test -; ranges from 0 to 1, or 0% to 100%)
  - If sensitivity = 0.8 (80%), 1-sensitivity = 0.2 (20% false negatives)
- Specificity: true negatives (proportion of individuals without the disease who test -; ranges from 0 to 1, or from 0% to 100%)
- 1-Specificity: false positives (proportion of individuals without the disease who test +; ranges from 0 to 1, or 0% to 100%)
  - If specificity = 0.92 (92%), 1-specificity = 0.08 (8% false positives)

Receiver Operating Curve (ROC) and Areas under the Curve (AUC)

- Plots sensitivity of the test (true + rate, TPR) on Y axis, from 0 to 1 vs. 1-specificity (false + rate, FPR) on X axis, from 0 to 1 at different test cutoffs
- Perfect Classification:
  AUC=1 (area of a square with sides=1)
- Random guess:
  AUC=0.5 (area of a triangle with base and height=1) (see **B**)
- AUC between 0.5 and 1:
  Test is **Better** than a random guess (see **A** and **C**)
- AUC between 0 and 0.5:
  Test is **Worse** than a random guess (see **D**)
- AUC also has a 95% CI
  - e.g., 0.78 (0.69-0.87)



ROCs of 3 tests with AUCs better than a random guess (AUC 0.5-1)



Likelihood Ratios (LR) and Positive/Negative Predictive Values (PPV/NPV) can be easily derived from Sensitivity and Specificity

Likelihood ratios: does the test *usefully* change the probability (likelihood) of a disease or condition?

Positive (+) likelihood ratio= true+/false + =sensitivity/1-specificity.

- The higher the + likelihood ratio, the more confident we are that the patient has the condition if the test is +. **+ LR** can approach ∞.

Negative (-) likelihood ratio = false-/true - = 1- sensitivity/ specificity.

- The lower the – likelihood ratios, the more confident we are that the patient does not have the condition if the test is -. **– LR** can approach 0.

Example: Use of + and - likelihood ratios

- Your patient with COPD has an acute onset of worsening dyspnea. He had arthroscopic knee surgery 2 weeks ago. There is no leg swelling or leg pain, hemoptysis, personal or family history PE or DVT, or malignancy. You clinically assess the odds of him having a PE as 50:50 (1:1), or equally likely that he had a PE as that he did not have a PE (eg, COPD exacerbation).
- If ordered and performed, how would the results of a pulmonary artery CT angiogram (CTA) change your estimated likelihood of PE in this patient? In other words, how good would a CTA be in helping you diagnose or exclude a PE in this patient?

Example, cont'd

Literature (Annals Internal Medicine 136: 286-287, 2002):

Pulmonary CTA and pulmonary angiography (gold standard) were performed in 250 patients with possible PE.

50 (20%) of the patients had PE on pulmonary angiography. 200 (80%) had no PE on angiography.

Results:

	CTA+	CTA-
PE on pulm angio (n=50)	35	15
No PE on pulm angio (n=200)	2	198

Example 1, continued

Likelihood ratio (LR) calculations for CTA:

CTA sensitivity (true +)= 35/50 (.70) , or 70%

1-sensitivity (false -)= 15/50 (.30), or 30%

CTA specificity (true -)= 198/200 (.99), or 99%

1-specificity (false +)=2/200 (.01), or 1%

+LR = sensitivity/1-specificity = true+/false+ = .70/.01=70
(PE 70 x as likely as before test). 1:1→70:1

-LR = 1-sensitivity /specificity= false-/true- = .30/.99=.30
(PE .30 x as likely as before test) 1:1 →0.3:1

Example 1, continued

PPV and NPV calculations for CTA:

CTA sensitivity (true +)= 35/50 (.70) , or 70%

1-sensitivity (false -)= 15/50 (.30), or 30%

CTA specificity (true -)= 198/200 (.99), or 99%

1-specificity (false +)=2/200 (.01), or 1%

PPV for CTA= true+/(true+ plus false+)= 35/37= 95%

NPV for CTA= true-/(true- plus false -)= 198/213= 93%

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

- 2-by-2 tables (Chi square, Fisher exact, others)
- Odds ratios or hazard ratios

- Sensitivity, Specificity and Receiver Operator Curves/AUCs
- Likelihood ratios and Positive/Negative predictive Values

- **Tests to Assess Statistical Significance (p values)**

- **Non-inferiority study designs**

Ann Int Med 2009: 150: JC6-16

What Test(s) to Use:

Continuous variable, normally distributed: Use student's t test

- Use a *paired* t if each subject is his/her own control **1**
  - Usually cross-over design
- Use an *unpaired* t (group t) if there are two groups **2**
  - Usually where group assignment is random

If data are <u>not</u> normally distributed:

If the variable is **continuous**, such as age or PaO_2?

- Use Wilcoxon's sign rank test for *paired* data **3**
- Use Mann Whitney U test for *unpaired* data **4**

If the variable is categorical, such as gender or smoking

- Use Fisher's exact test, **5**

If there >2 study groups:

Use analysis of variance (ANOVA) or covariance (ANCOVA) **6**

What Tests to Use:
Correlations (r) between variables

If the variables are normally distributed:

Use Pearson's test **7**:

Pearson's r ranges from -1 to +1.

$r \cong 0$ indicates no correlation.

If the variables are not normally distributed:

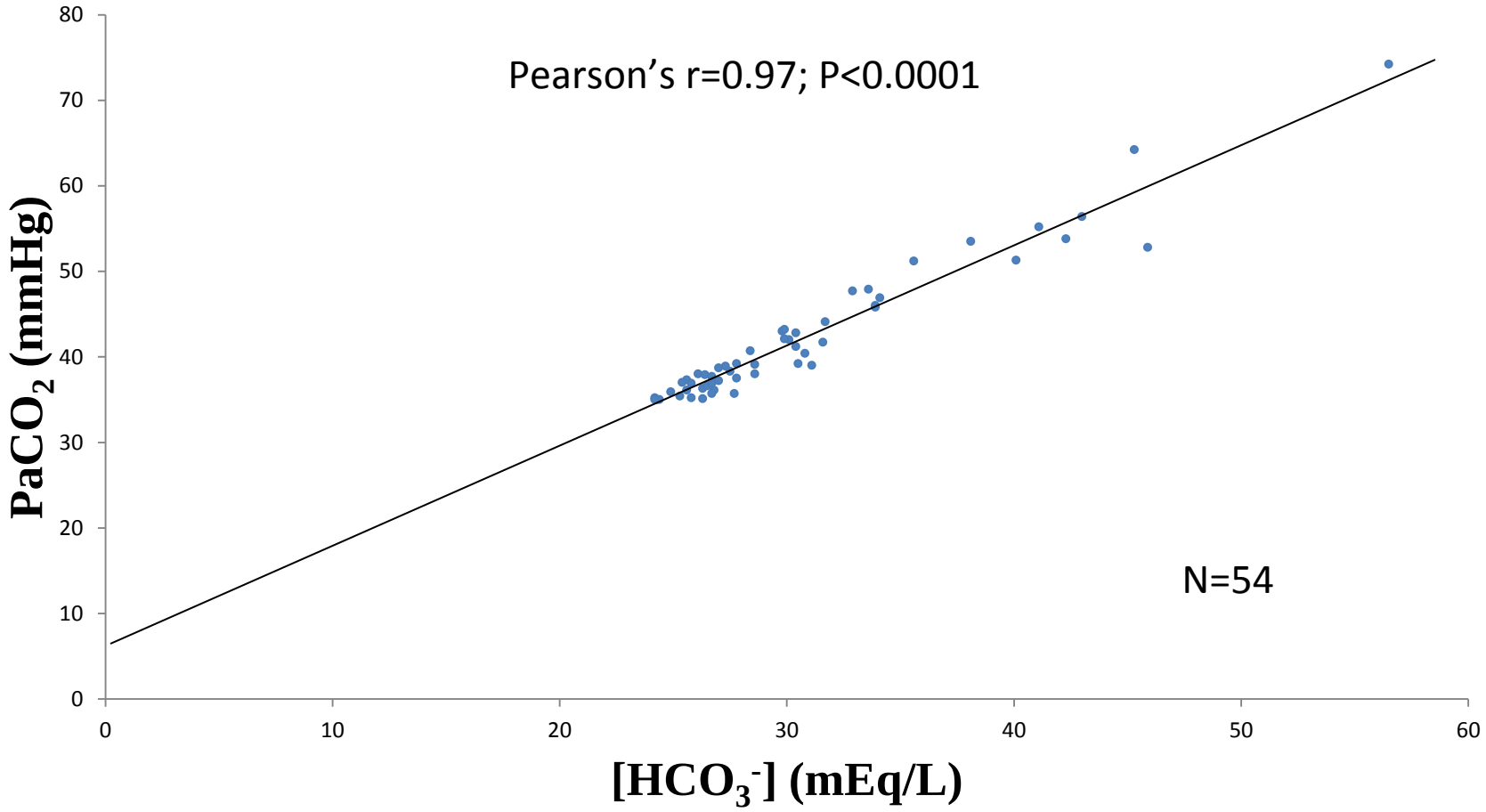
Use Spearman's test **8**:

Spearman's r ranges from -1 to +1

$r \cong 0$ indicates no correlation.

P values depend both on r and N. $P < 0.05$ usually used.

METABOLIC ALKALOSIS (Feldman and Alvarez)



Free Online Resources for Common Tests of Statistical Significance

TEST	WEBSITE
Paired t	http://graphpad.com/quickcalcs
Unpaired t	http://graphpad.com.quickcalcs
Fisher exact	http://graphpad.com/quickcalcs
Mann Whitney/Wilcoxon/ANOVA/etc.	http://vassarstats.net/

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

- 2-by-2 tables (Chi square, Fisher exact, others)
- Odds ratios or hazard ratios

- Sensitivity, Specificity and Receiver Operator Curves/AUCs
- Likelihood ratios and Positive/Negative predictive Values

- Tests to Assess Statistical Significance (p values)

- **Non-inferiority study designs**

Ann Int Med 2009: 150: JC6-16

New Treatments and Trials

A New Treatment Can Truly Be:

- Better (Superior)
- Equal
- Worse (Inferior)
  - than the usual treatment

A Trial Can Test Whether New is:

- Better or Worse
  - superiority trial
  - inferiority trial
- Not better (non-superiority trial)*
- Not worse (non-inferiority trial)

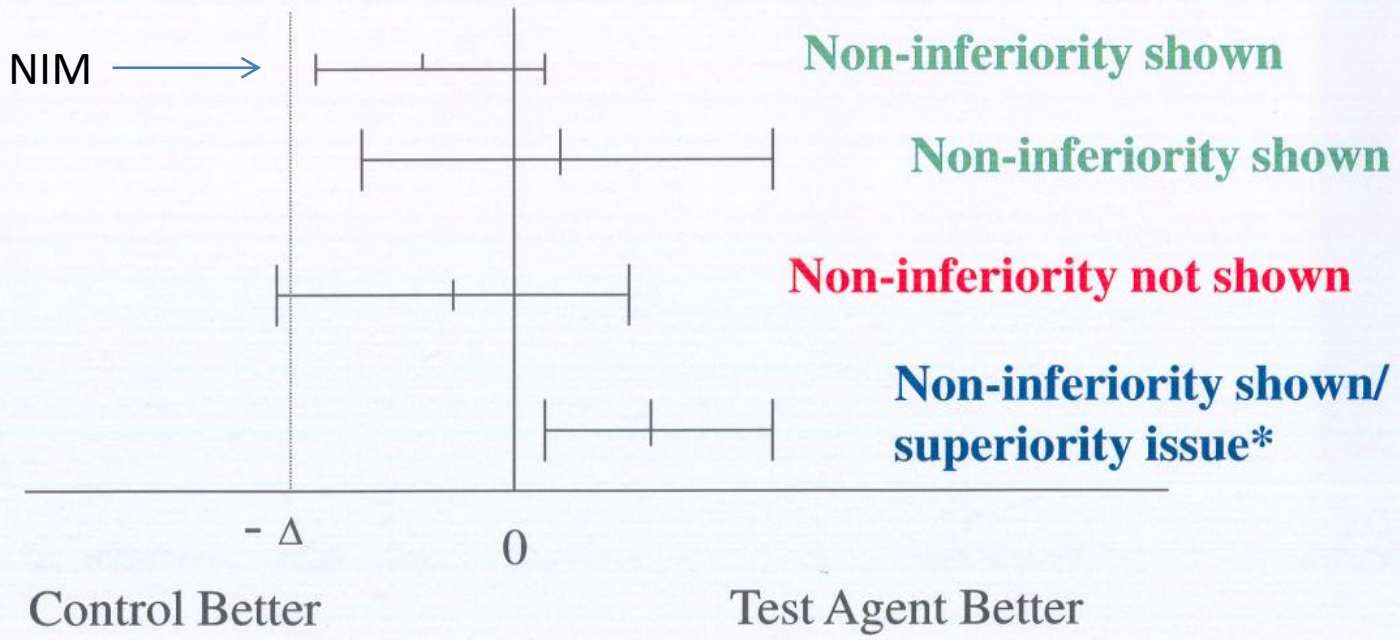
* rarely done

Non-inferiority trials

- Non-inferiority trials are intended to show that the effect of a **new** treatment is not worse than that of an **active** control by more than a specified amount.
- The non-inferiority margin (NIM) is chosen by the investigators before the study (*a priori*) and can be somewhat arbitrary.
- Study endpoints in non-inferiority trials can be efficacy or safety parameters or a combination of the two.
- Study design may include 3 arms with placebo group (preferred) or 2 arms with only new and usual treatments (much less ideal, since no internal validation that new treatment is better than placebo)
- *Delta* (Δ) is the measured difference (*best estimate of the true difference*) between the two active treatments. Δ will have a 95% CI.
  - **Example**: Δ = -4% (95% CI, -9% to +1%)

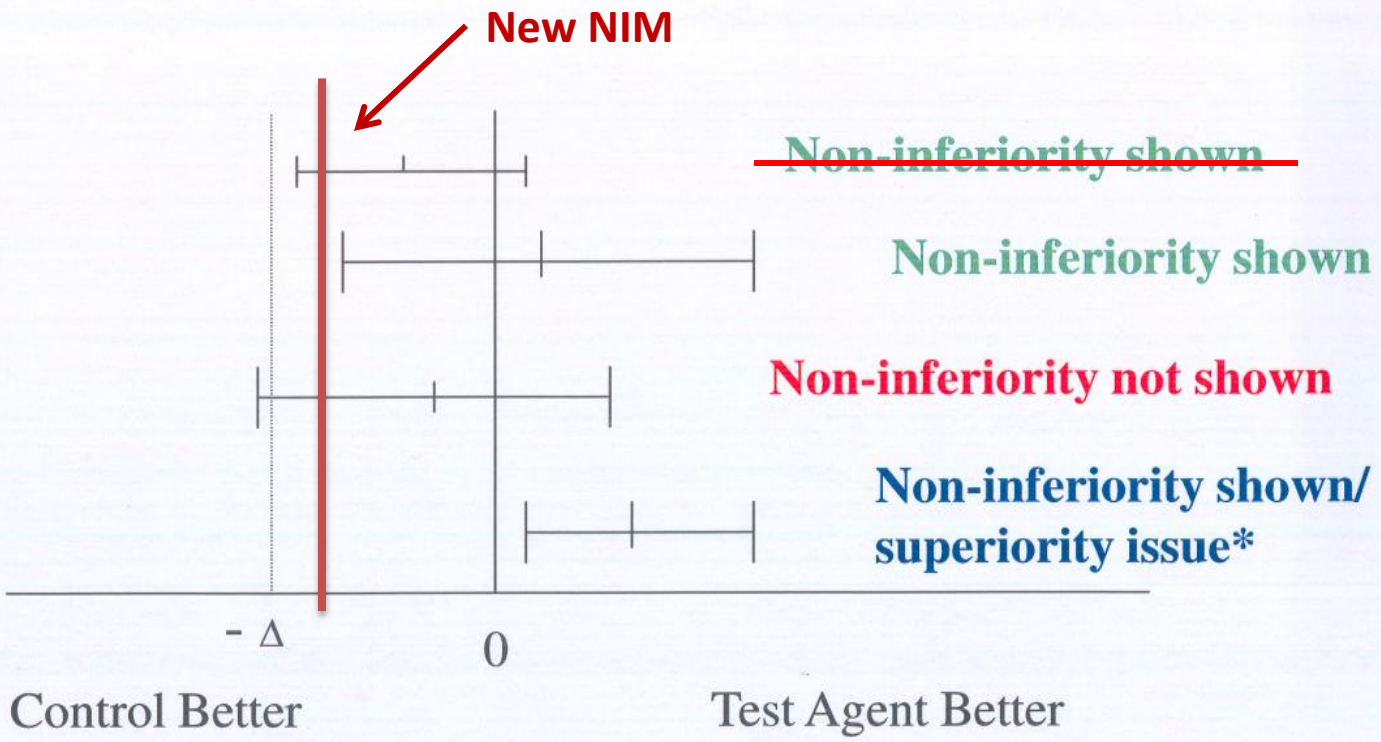
Inference for Non-Inferiority

Delta Limits (95%) and Confidence Intervals



Inference for Non-Inferiority

Delta Limits (95%) and Confidence Intervals



Non-Inferiority Trial using Hazard Ratios: EINSTEIN-PE study

- Non-inferiority trial of rivaroxaban (Xarelto) versus warfarin or acenocoumarol in PE
- "Assuming equal efficacy of the two study treatments, we determined that 88 events would provide 90% power (1-β) to show that rivaroxaban was non-inferior to standard therapy, *using a margin of 2.0 for the upper limit of the 95% confidence interval for the observed hazard ratio*, with a two-sided α level of 0.05."
- Results: Rivaroxaban had 50 events vs. 44 in standard therapy group, with HR of 1.12 (0.75-1.68).
- Note: 1.68 is < 2.0.
- Authors' conclusion: Rivaroxaban is noninferior to vit K antagonist in PE.
  - What if NIM had been set *a priori* at 1.6 instead of 2.0?