AN OVERVIEW OF BIOSTATISTICS

UNC CENTER FOR AIDS RESEARCH

August 23, 2013

Katie Mollan

Michael Hudgens

UNC CFAR
Center for AIDS Research

CFAR BIOS Core

Michael Hudgens, PhD
Core Director

Katie Mollan, MS
Biostatistician

Bios PhD Student, GRA

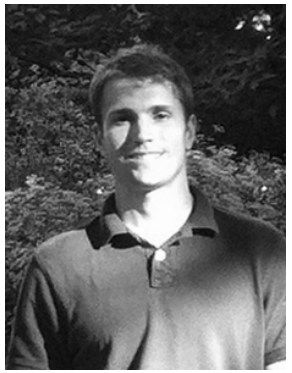
Stephen R. Cole, PhD
Professor of Epidemiology
CNICS Epi and Bios Core Director

Sonia Napravnik, PhD
Assistant Professor of Medicine

CFAR BIOS Core

Michael Hudgens, PhD
Core Director

Katie Mollan, MS
Biostatistician



Joe Rigdon
Bios PhD Student, GRA

Stephen R. Cole, PhD
Professor of Epidemiology
CNICS Epi and Bios Core Director

Sonia Napravnik, PhD
Assistant Professor of Medicine

An Overview of Biostatistics

- Know Your Data
- Statistical Graphics
- Statistical Inference
- Sample Size/Power
- Resources/References
- Pitfalls along the way

An Overview of Biostatistics

- Not an exhaustive list or catalog of which statistical methods to employ in various settings
- (Almost) no math/formulas
- No software training

- General principles and guidelines
- Mistakes to avoid
- Additional resources

“Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write.”
-H.G. Wells

KNOW YOUR DATA

Moving from data to results

A) Merging databases to make analysis datasets

- Major part of analyzing big studies & clinical trials

B) Describe the data

- N, mean, median, Q1, Q3, standard deviation, frequency (%)
- Review data, visualize with graphics

C) Clean & query the data as needed

- Intersection of statistics and data management

D) Statistical analysis

E) Interpret and report results

Look at your data

- Review data records for oddities
  - Out of range values
  - Missing values
  - Pitfall: logical inconsistencies (e.g. adults getting shorter, dates out of order, dead but still on-study)
- Find out why data are missing
- Large studies
  - Review a subset of records, use frequency listings to review data combinations and missing data patterns
  - Include logic checks in your database

Dataset Example (Rectangle, i.e. Matrix)

ID	AGE	SEX	HDL	CHOL	CD4PCT
1	35	M	59	178	54
2	42	M	65	220	35
3	68	M	45	213	187
4	34	F	55	195	16
5	31	M	60	187	10
6		M	38	205	42
7	46		55	180	37
8	64	F			48
9	50	F	45	195	57
10	14	M	198	46	32

Hypothetical (not real) data set

Dataset Example (Potential Problems)

ID	AGE	SEX	HDL	CHOL	CD4PCT
1	35	M	59	178	54
2	42	M	65	220	35
3	68	M	45	213	187
4	34	F	55	195	16
5	31	M	60	187	10
6		M	38	205	42
7	46		55	180	37
8	64	F			48
9	50	F	45	195	57
10	14	M	198	46	32

Dataset Example Continued..

Corrections were made.

Now it's looking better..

But do you notice anything odd?

ID	AGE	SEX	HDL	CHOL	CD4PCT
1	35	M	59	178	54
2	42	M	65	220	35
3	68	M	45	213	32
4	34	F	55	195	16
4	34	F	55	195	16
5	31	M	60	187	10
6	53	M	38	205	42
7	46	F	55	180	37
8	64	F			48
9	50	F	45	195	57
10	41	M	46	198	32

Dataset Example Continued..

The updated dataset has a duplicate row!

ID	AGE	SEX	HDL	CHOL	CD4PCT
1	35	M	59	178	54
2	42	M	65	220	35
3	68	M	45	213	32
4	34	F	55	195	16
4	34	F	55	195	16
5	31	M	60	187	10
6	53	M	38	205	42
7	46	F	55	180	37
8	64	F			48
9	50	F	45	195	57
10	41	M	46	198	32

Data Example 2

Imagine this dataset continues for n=50 patient IDs…

Notice anything strange?

ID	BIOMARKER	RESULT
1	IFNg	1.50
1	TGFa	2.46
1	IL6	5.87
1	IL7	1.50
1	IL10	5.12
2	IFNg	2.31
2	TGFa	10.23
2	IL6	5.15
2	IL7	1.50
2	IL10	6.24
3	IFNg	9.60
3	TGFa	1.50
3	IL6	1.50
3	IL7	8.42
3	IL10	7.46

Data Example 2

Some results are censored at the limit of detection.

This is important to know for choosing an analysis method

(Probably cannot assume normality)

ID	BIOMARKER	RESULT	CENSOR
1	IFNg	1.50	L
1	TGFa	2.46	
1	IL6	5.87	
1	IL7	1.50	L
1	IL10	5.12	
2	IFNg	2.31	
2	TGFa	10.23	
2	IL6	5.15	
2	IL7	1.50	L
2	IL10	6.24	
3	IFNg	9.60	
3	TGFa	1.50	L
3	IL6	1.50	L
3	IL7	8.42	
3	IL10	7.46	

Know your data summary

- Look at records from the original data, intermediate data sets, and analysis data
- Use cross-frequency listings, plots, range checks, and logic checks to “look at the data”
- Be careful with coded variables, be sure you know what 0/1 represent, 1, 2, 3.. and so on

Know your data. Pitfalls + Scandal

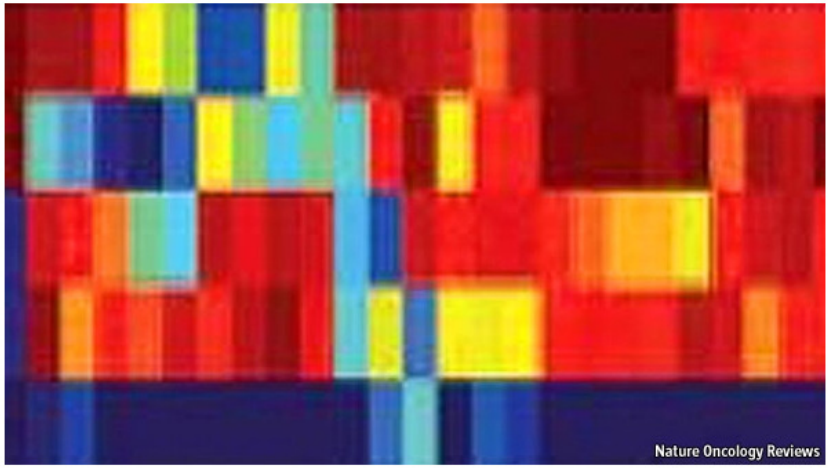
Misconduct in science

An array of errors

The Economist

Investigations into a case of alleged scientific misconduct have revealed numerous holes in the oversight of science and scientific publishing

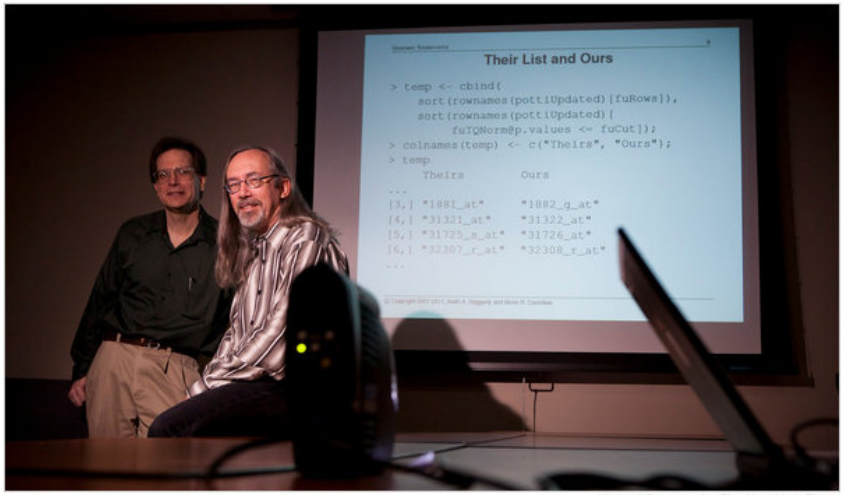
Sep 10th 2011 | From the print edition



ANIL POTTI, Joseph Nevins and their colleagues at Duke University in Durham, North Carolina, garnered widespread attention in 2006. They reported in the *New England Journal of Medicine* that they could predict the course of a patient's lung cancer using devices called expression arrays, which log the activity patterns of thousands of genes in a sample of tissue as a colourful picture (see above). A few months later, they wrote in *Nature Medicine* that they had developed a similar technique which used gene expression in laboratory cultures of cancer cells, known as cell lines, to predict which chemotherapy would be most effective for an individual patient suffering from lung, breast or ovarian cancer.

The New York Times

How Bright Promise in Cancer Testing Fell Apart



Michael Stravato for The New York Times

Keith Baggerly, left, and Kevin Coombes, statisticians at M. D. Anderson Cancer Center, found flaws in research on tumors.

By GINA KOLATA
Published: July 7, 2011

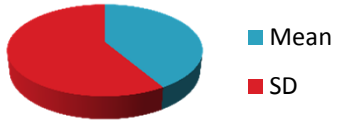
Errors.. "Some seemed careless — moving a row or a column over by one in a giant spreadsheet — while others seemed inexplicable." -NYTimes

Useful Reference:

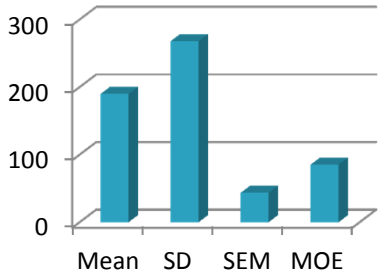
Pocock, Stuart J., Thomas G. Travison, and Lisa M. Wruck. "Figures in clinical trial reports: current practice & scope for improvement." *Trials* 8.1 (2007): 36.

STATISTICAL GRAPHICS

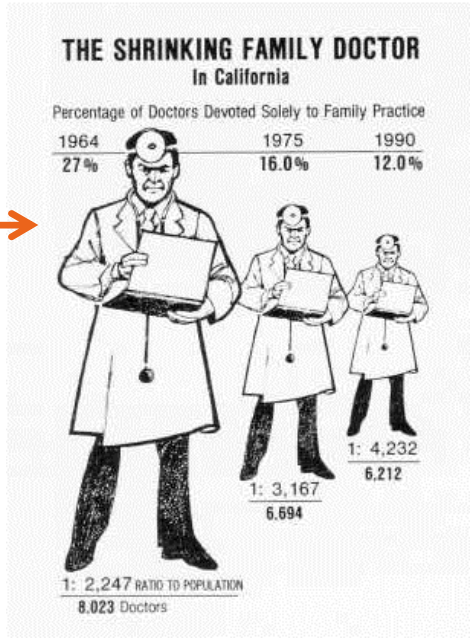
Pitfalls: What not to plot



Friends don't let friends plot imaginary dimensions



- Don't use a 3D chart unless the 3rd dimension displays another variable
- Don't use 2 dimensions (area) to convey one-dimensional data
- Pie charts are rarely ideal, bar charts provide an axis for orientation

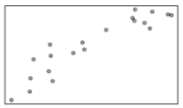


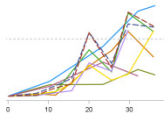
-Tufte, 1983 pg. 69

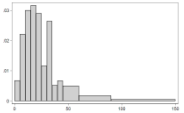
How to Lie with Statistics – Huff 1954
http://en.wikipedia.org/wiki/Misleading_graph

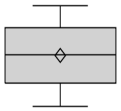
Statistical Graphics

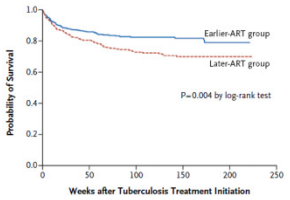
- Show the data, even a simple plot will do
  - Scatter Plot (association)
  - Spaghetti Plot (trajectories, paired data)
  - Histogram
- Summary Statistics
  - Box plot
  - Bar chart
- Estimation and inference
  - Kaplan-Meier Curves
  - Forest Plot

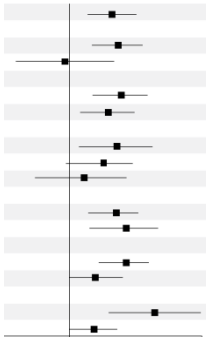






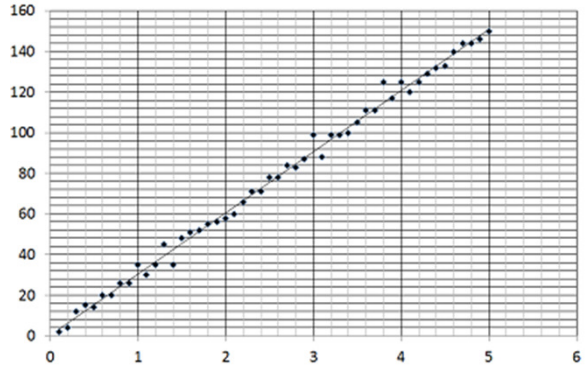






Aim for high data-ink ratio

$$\text{data-ink ratio} = \frac{\text{‘ink’ used to display the data}}{\text{total ‘ink’ used to display the graphic}}$$

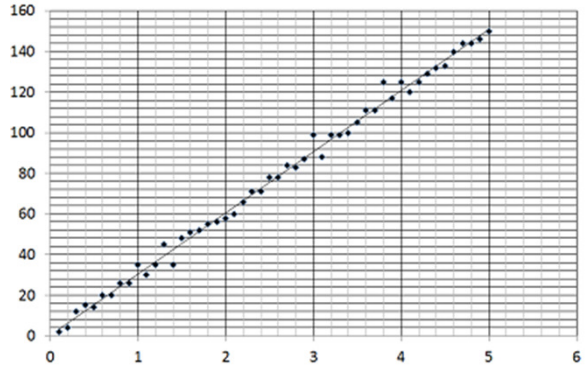


Low ratio
Data are lost in a sea of gridlines and labels

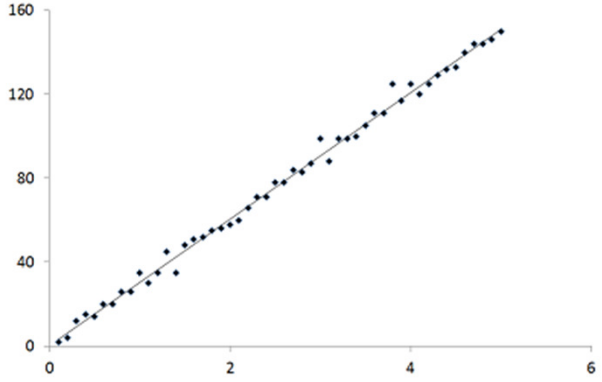
http://en.wikipedia.org/wiki/Misleading_graph
http://www.davidgiard.com/2011/05/12/DataVisualizationPart5DataInk.aspx

Aim for high data-ink ratio

$$\text{data-ink ratio} = \frac{\text{‘ink’ used to display the data}}{\text{total ‘ink’ used to display the graphic}}$$



Low ratio
Data are lost in a sea of gridlines and labels

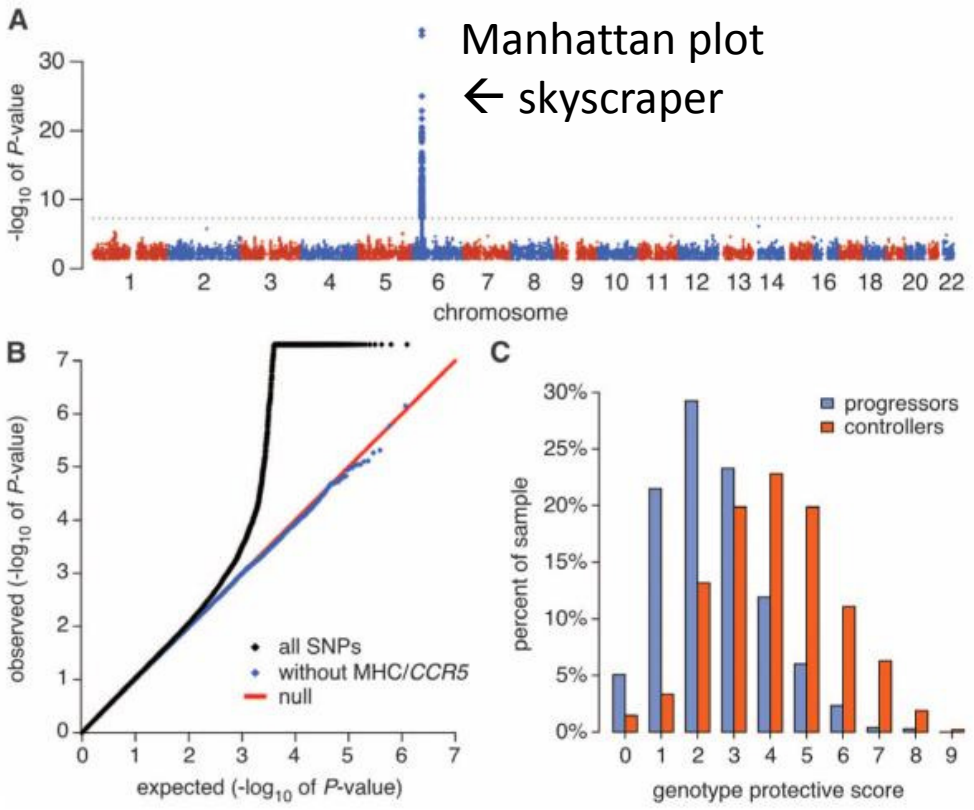


High ratio ☺
Ah.. There we go

http://en.wikipedia.org/wiki/Misleading_graph
http://www.davidgiard.com/2011/05/12/DataVisualizationPart5DataInk.aspx

Graphics rules of thumb

- Label the axes with reasonable size font
- Always start axis at zero when applicable
- Be thoughtful and fair with graph dimensions
- Use axis breaks sparingly
- Use gridlines and boxes sparingly



Pereyra, Florencia, et al. "The major genetic determinants of HIV-1 control affect HLA class I peptide presentation."
***Science (New York, NY)* 330.6010 (2010): 1551.**

Sophisticated graphic from *Science*

Scatter Plot of Pharmacy Refill Adherence Levels at 12 Months after Starting cART for Patients with and without Virologic Failure

Bisson GP, et al. (2008) Pharmacy Refill Adherence Compared with CD4 Count Changes for Monitoring HIV-Infected Adults on Antiretroviral Therapy. **PLoS Med 5(5): e109. doi:10.1371/journal.pmed.0050109**
http://www.plosmedicine.org/article/info:doi/10.1371/journal.pmed.0050109

PLOS | MEDICINE

Scatter Plot of Quadratic Association

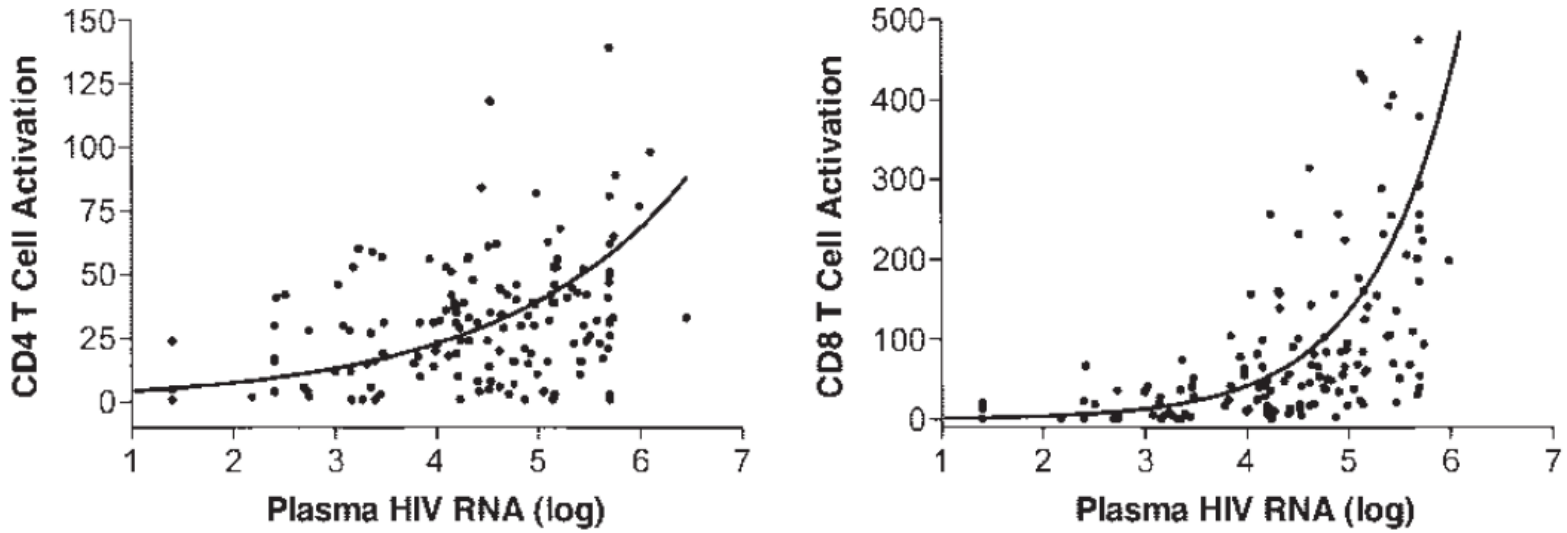
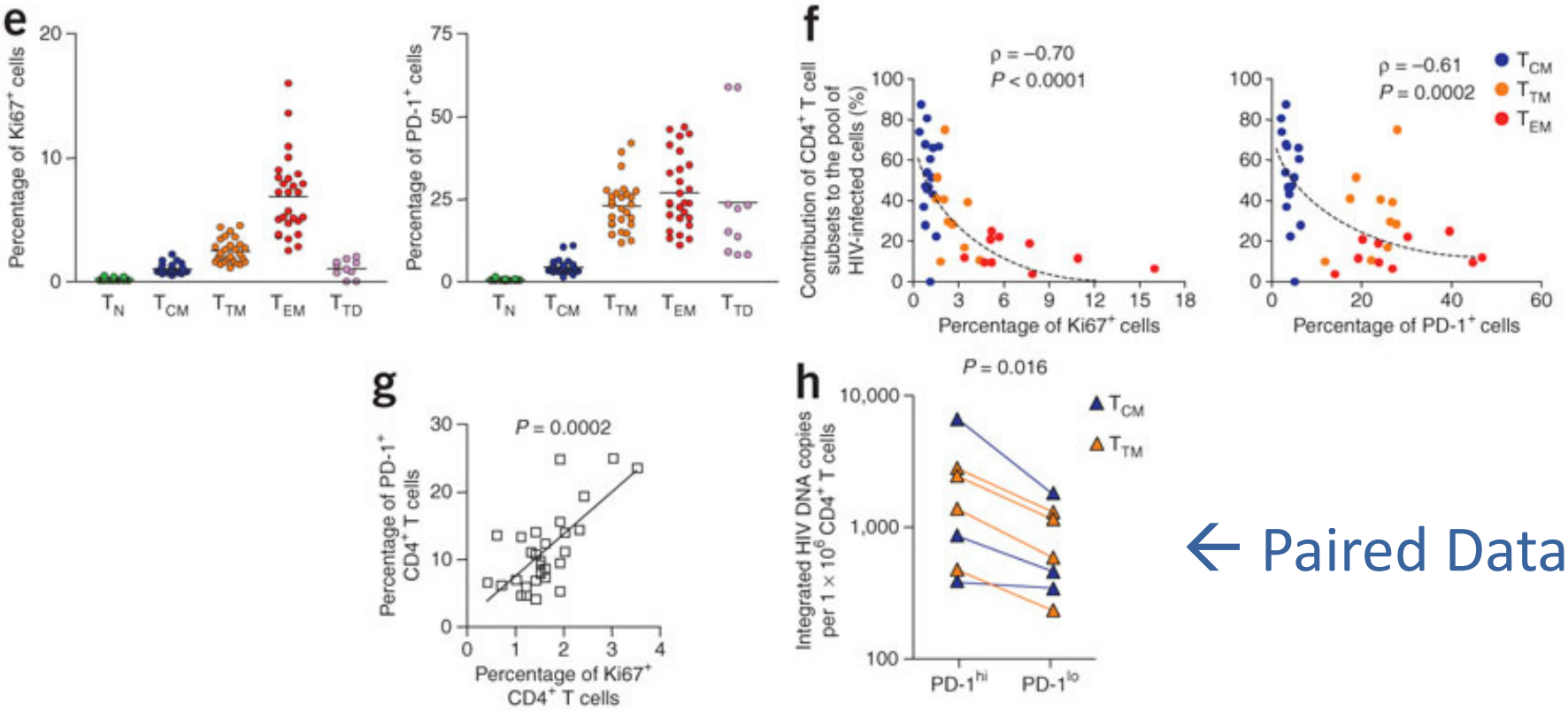


Figure 1. The relationship between T-cell activation and plasma HIV RNA levels (log_{10} transformed) in 153 individuals recently diagnosed with HIV infection. A smooth line was generated by linear regression with quadratic equations.

Deeks, Steven G., et al. "Immune activation set point during early HIV infection predicts subsequent CD4+ T-cell changes independent of viral load." ***Blood*** **104.4 (2004): 942-947.**

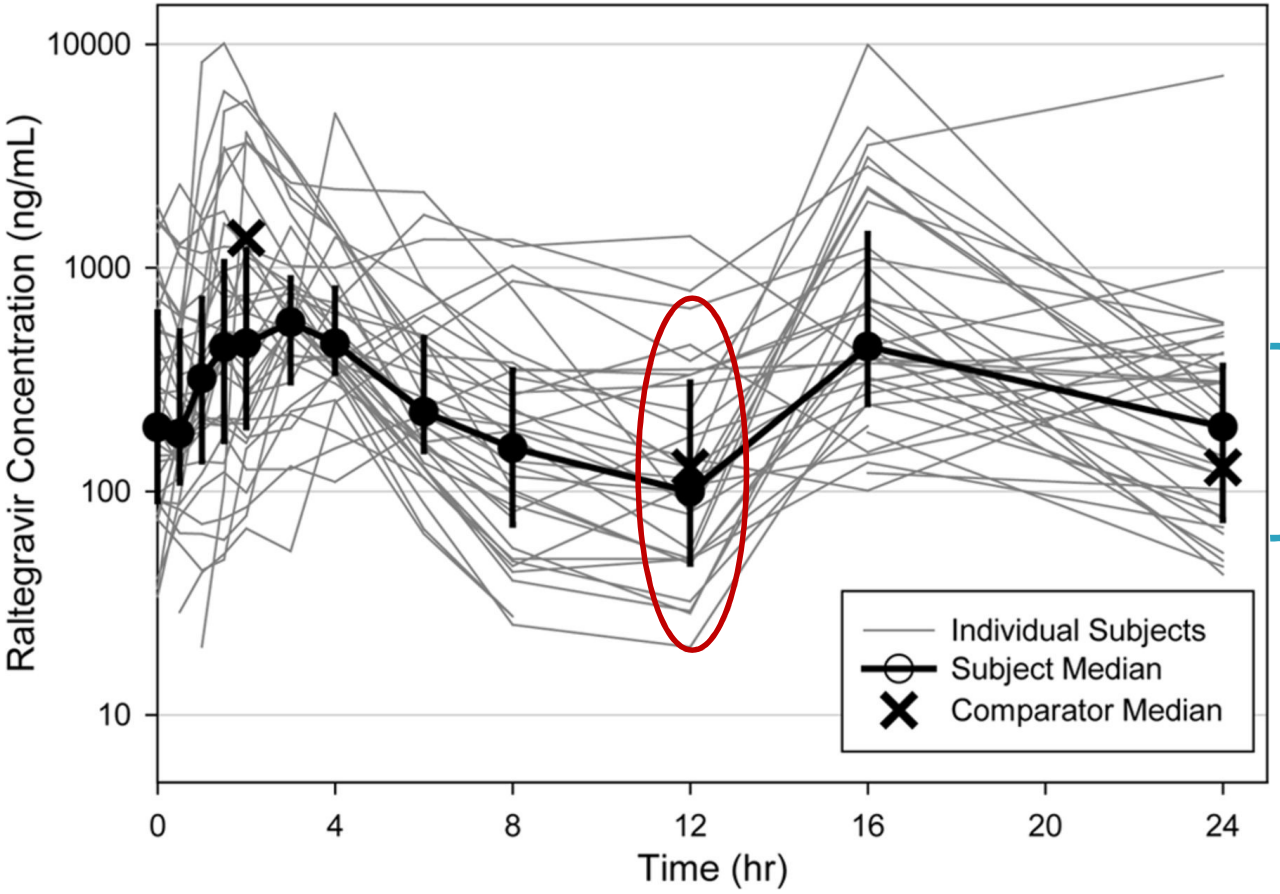
Nice use of color



← Paired Data

Chomont, Nicolas, et al. "HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation."
***Nature medicine* 15.8 (2009): 893-900.**

Individual subject concentration-time profiles over the 24-h study period for African-American REAL cohort participants (n = 38)



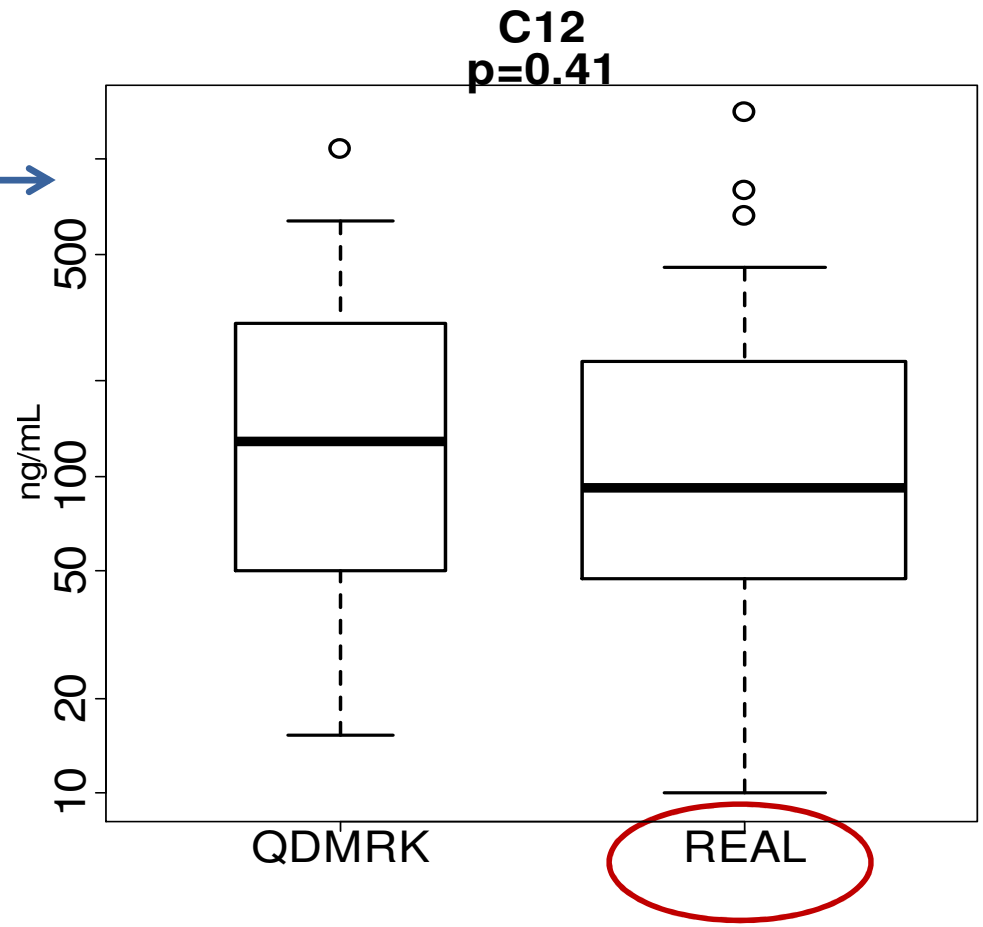
Interquartile Range 25^{th} to 75^{th} percentile

Wohl D A et al. Antimicrob. Agents Chemother. 2013;57:784-788

Antimicrobial Agents and Chemotherapy

Box Plot REAL Study Example

Plotted Log10 scale



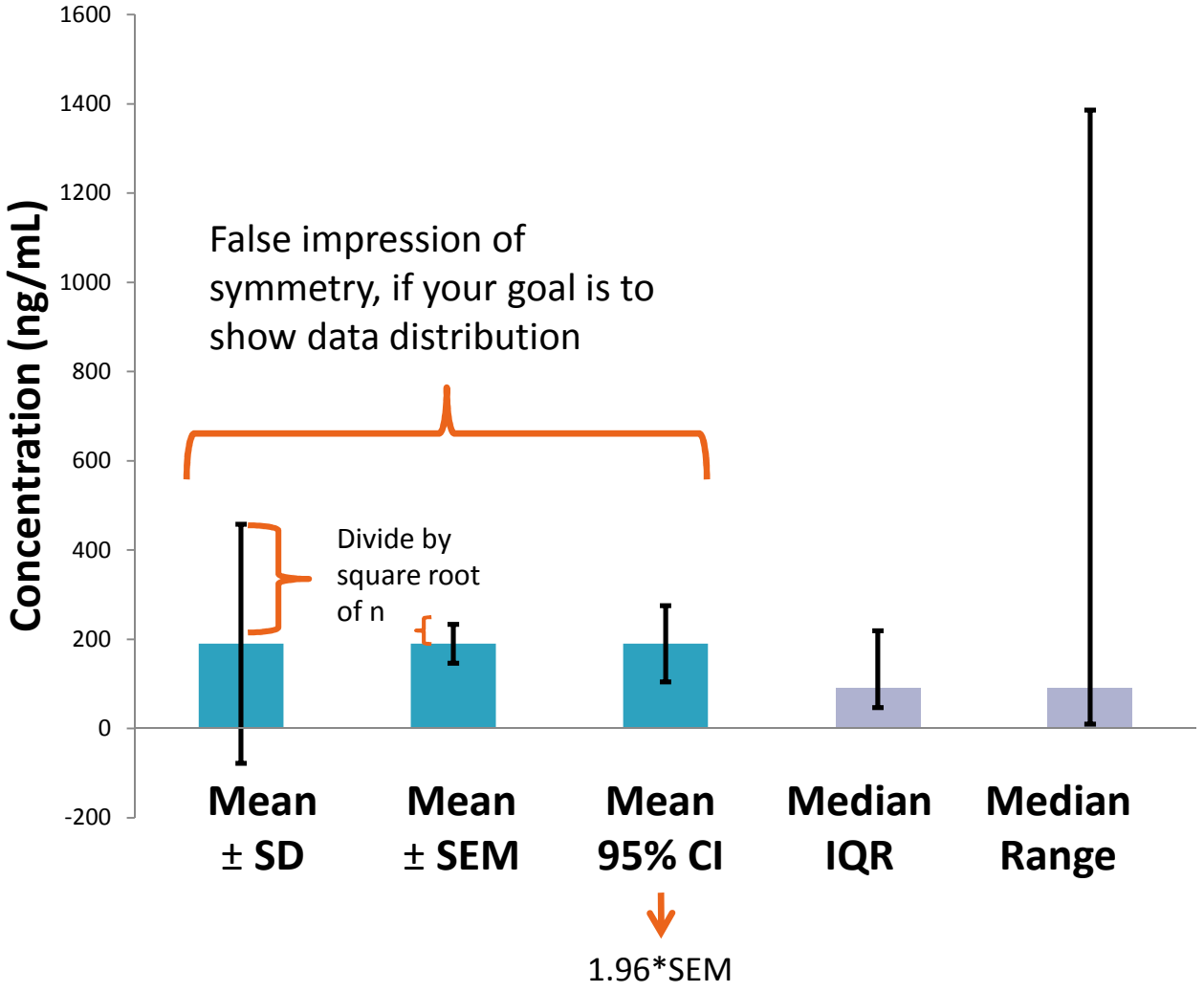
Wilcoxon Rank-Sum nonparametric test to compare 2 independent groups (QDMRK vs. REAL)

Section of Figure 2:
Wohl D A et al. Antimicrob. Agents Chemother. 2013;57:784-788

Many Types of Error Bars

- Descriptive
  - ± 1 standard deviation (SD)
  - Interquartile Range (IQR), 25^{th} to 75^{th} percentile
  - Range, smallest to largest
- Statistical Inference
  - ± 1 standard error (SE or SEM),
  - Confidence interval (CI), such as 95% CI
- If the type of error bar is not specified, that is a problem

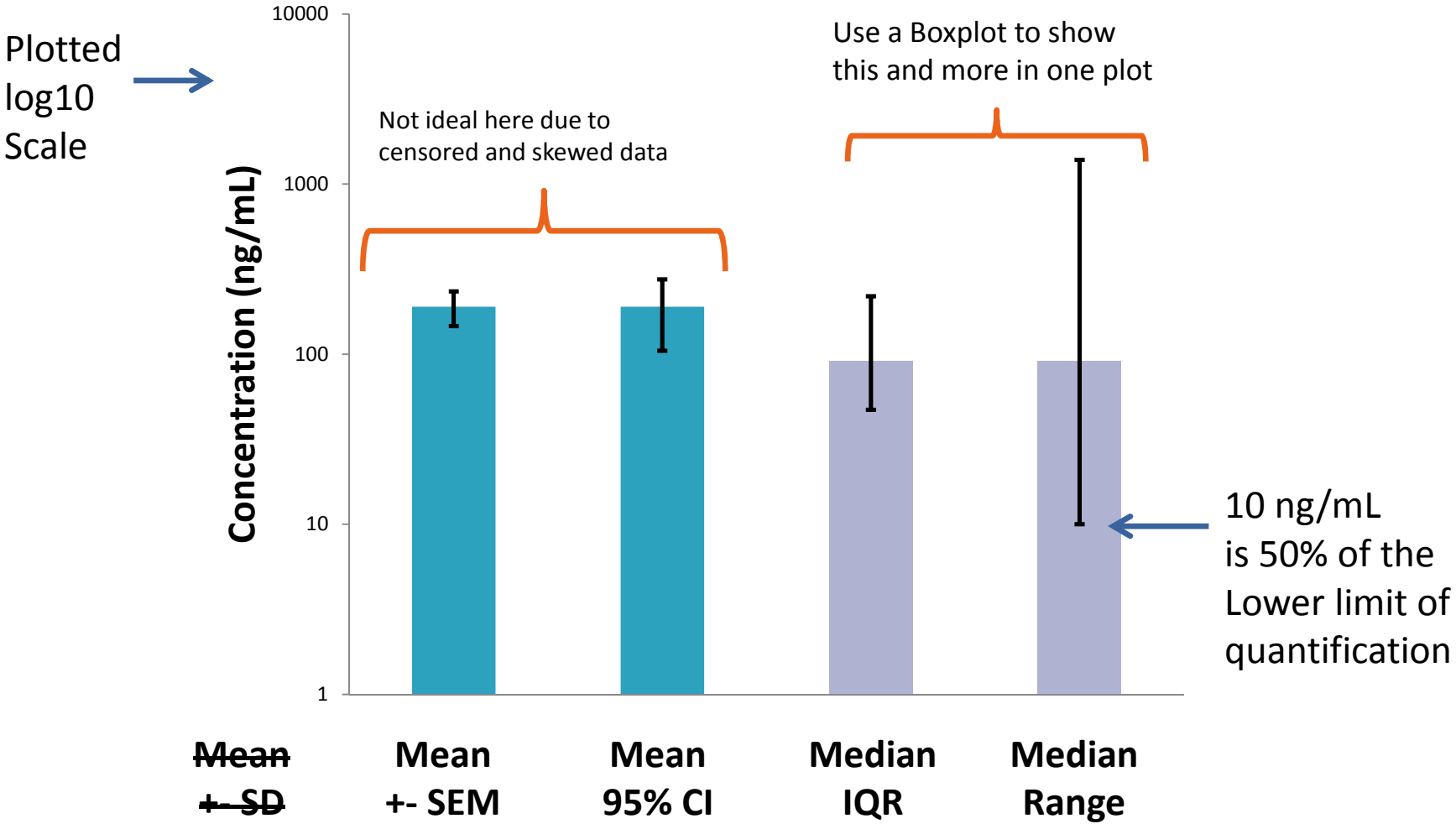
Error Bar Example
REAL Study C12 for Raltegravir (n=38)



Data from Table 2:
Wohl D A et al. Antimicrob. Agents Chemother. 2013;57:784-788

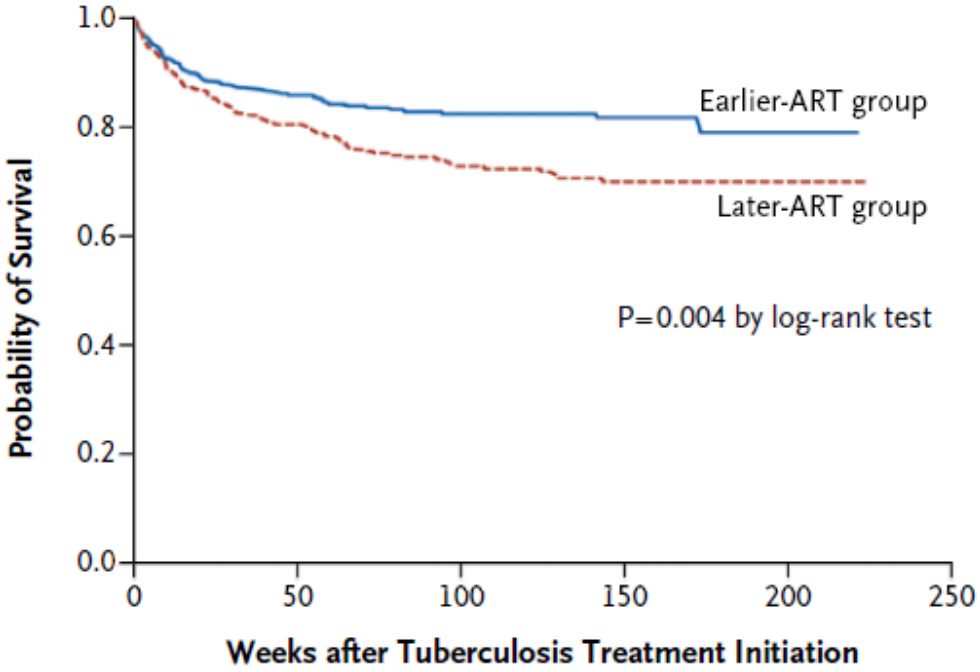
SEM: Standard Error of Mean

Change in Y-axis Scale



Data from Table 2:
Wohl D A et al. Antimicrob. Agents Chemother. 2013;57:784-788

Kaplan Meier Survival Curves



No. at Risk					
Earlier-ART group	332	278	192	101	4
Later-ART group	329	256	168	87	3
No. of Deaths					
Earlier-ART group	0	46	56	57	59
Later-ART group	0	63	85	90	90

Blanc, François-Xavier, et al. "Earlier versus later start of antiretroviral therapy in HIV-infected adults with tuberculosis." ***New England Journal of Medicine*** **365.16 (2011): 1471-1481.**

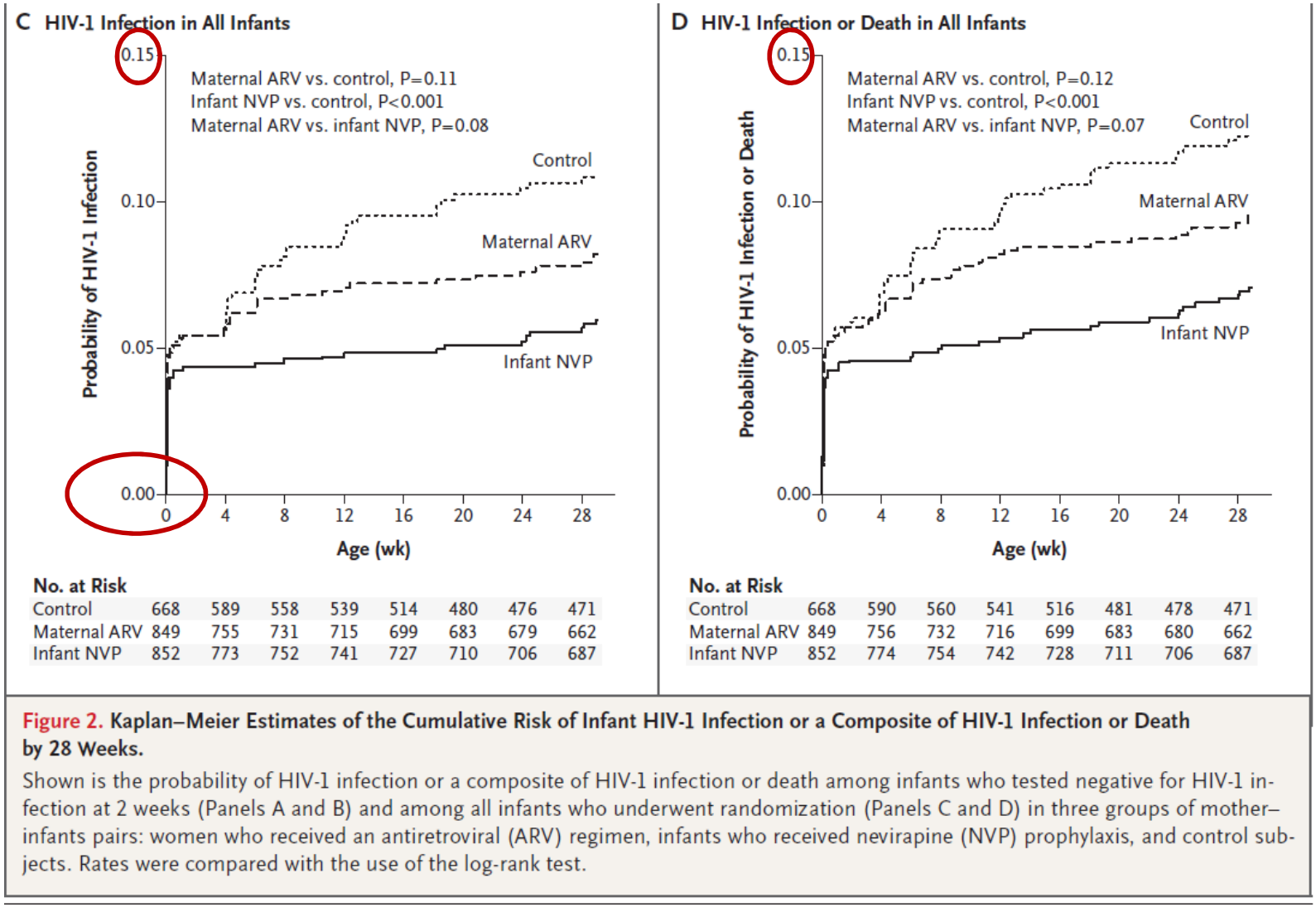


Figure 2. Kaplan–Meier Estimates of the Cumulative Risk of Infant HIV-1 Infection or a Composite of HIV-1 Infection or Death by 28 Weeks.

Shown is the probability of HIV-1 infection or a composite of HIV-1 infection or death among infants who tested negative for HIV-1 infection at 2 weeks (Panels A and B) and among all infants who underwent randomization (Panels C and D) in three groups of mother–infants pairs: women who received an antiretroviral (ARV) regimen, infants who received nevirapine (NVP) prophylaxis, and control subjects. Rates were compared with the use of the log-rank test.

Chasela, Charles S., et al. "Maternal or infant antiretroviral drugs to reduce HIV-1 transmission." ***New England Journal of Medicine*** **362.24 (2010): 2271-2281.**

KM Plot with Confidence Interval

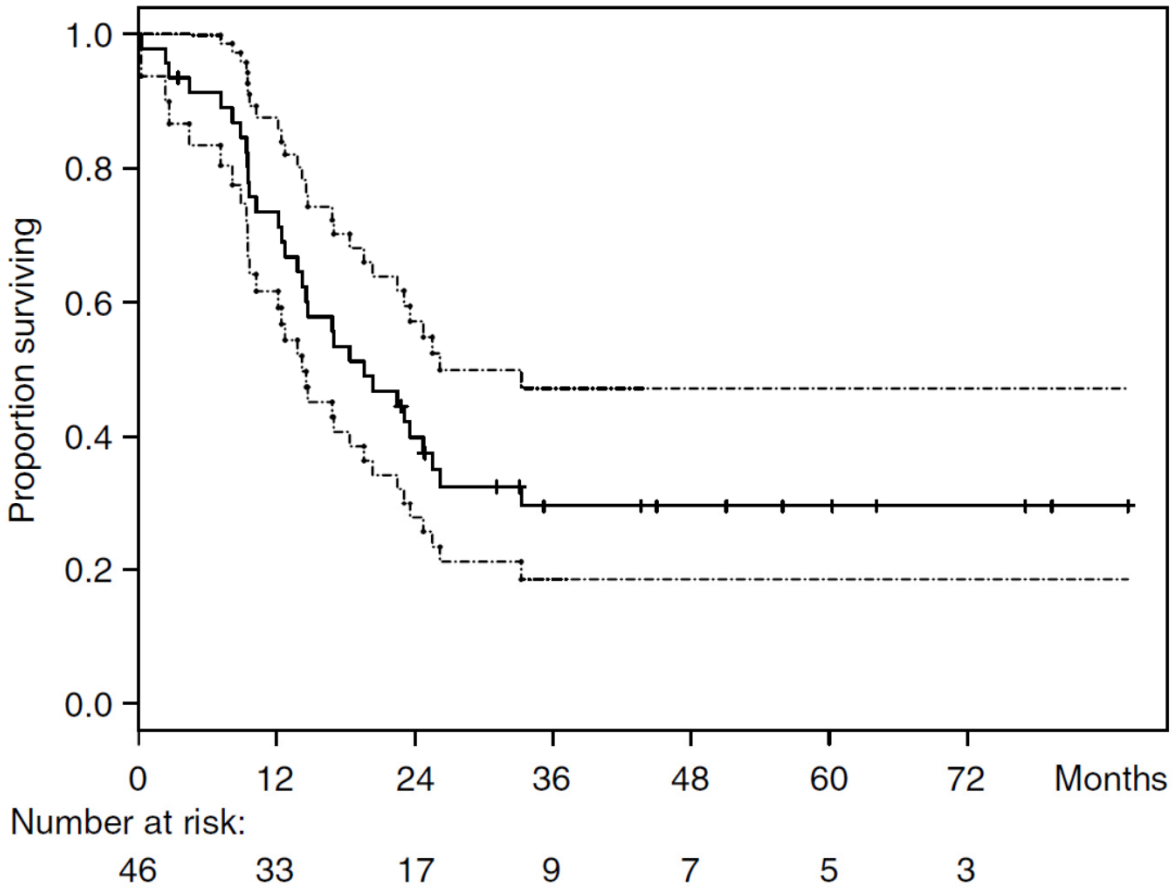
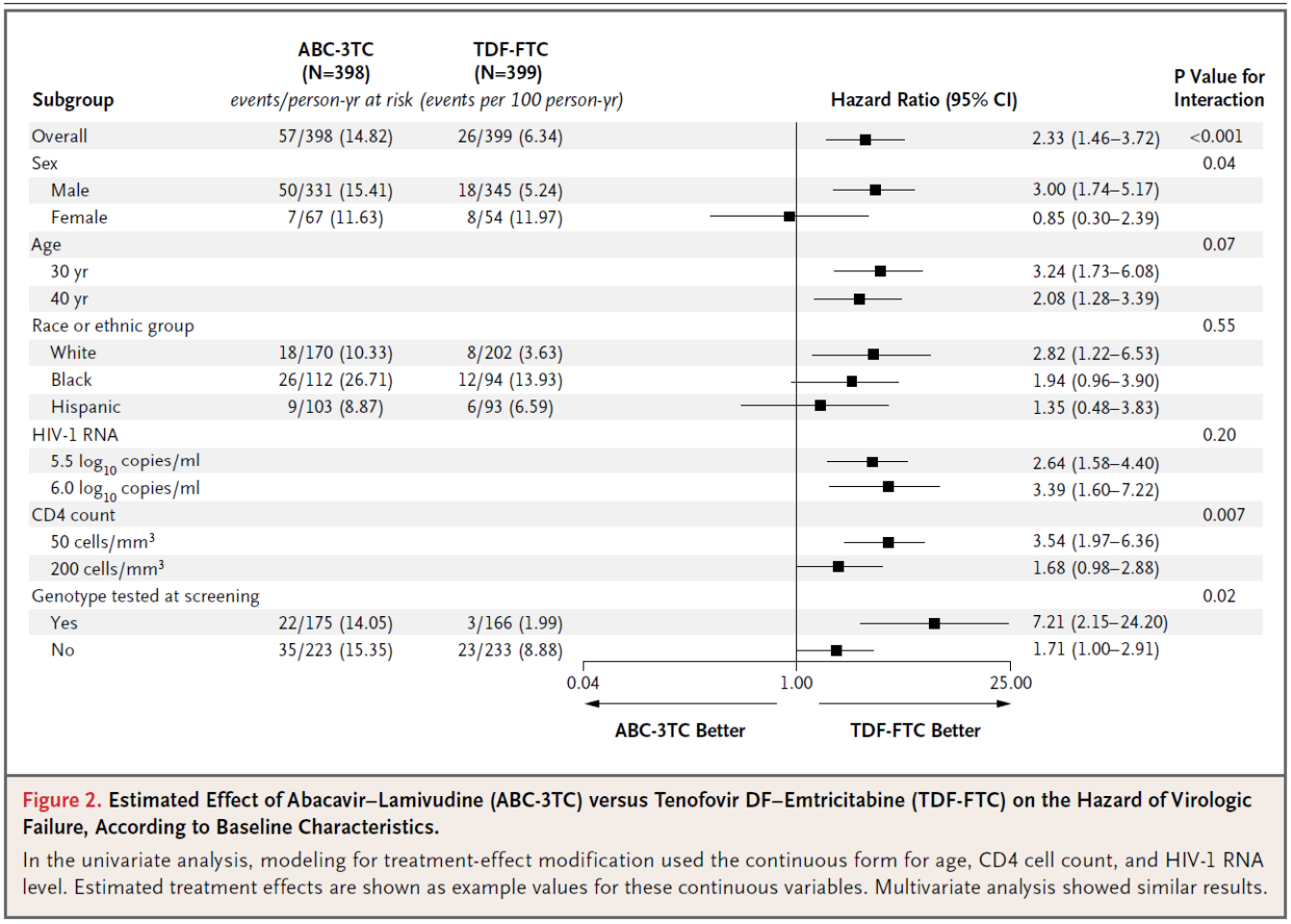


Figure 1 Kaplan–Meier curve for overall survival (interrupted line 95% confidence interval).

Leyvraz, S., et al. "Treatment of advanced soft-tissue sarcomas using a combined strategy of high-dose ifosfamide, high-dose doxorubicin and salvage therapies." ***British journal of cancer*** 95.10 (2006): 1342-1347.

Forest Plot



Subgroup	ABC-3TC (N=398) events/person-yr at risk	TDF-FTC (N=399) (events per 100 person-yr)	Hazard Ratio (95% CI)	P Value for Interaction
Overall	57/398 (14.82)	26/399 (6.34)	2.33 (1.46–3.72)	<0.001
Sex				0.04
Male	50/331 (15.41)	18/345 (5.24)	3.00 (1.74–5.17)	
Female	7/67 (11.63)	8/54 (11.97)	0.85 (0.30–2.39)	
Age				0.07
30 yr			3.24 (1.73–6.08)	
40 yr			2.08 (1.28–3.39)	
Race or ethnic group				0.55
White	18/170 (10.33)	8/202 (3.63)	2.82 (1.22–6.53)	
Black	26/112 (26.71)	12/94 (13.93)	1.94 (0.96–3.90)	
Hispanic	9/103 (8.87)	6/93 (6.59)	1.35 (0.48–3.83)	
HIV-1 RNA				0.20
5.5 $\log_{10}$ copies/ml			2.64 (1.58–4.40)	
6.0 $\log_{10}$ copies/ml			3.39 (1.60–7.22)	
CD4 count				0.007
50 cells/mm^3			3.54 (1.97–6.36)	
200 cells/mm^3			1.68 (0.98–2.88)	
Genotype tested at screening				0.02
Yes	22/175 (14.05)	3/166 (1.99)	7.21 (2.15–24.20)	
No	35/223 (15.35)	23/233 (8.88)	1.71 (1.00–2.91)	

Figure 2. Estimated Effect of Abacavir–Lamivudine (ABC-3TC) versus Tenofovir DF–Emtricitabine (TDF-FTC) on the Hazard of Virologic Failure, According to Baseline Characteristics.

In the univariate analysis, modeling for treatment-effect modification used the continuous form for age, CD4 cell count, and HIV-1 RNA level. Estimated treatment effects are shown as example values for these continuous variables. Multivariate analysis showed similar results.

Sax, Paul E., et al. "Abacavir–lamivudine versus tenofovir–emtricitabine for initial HIV-1 therapy." ***New England Journal of Medicine* 361.23 (2009): 2230-2240.**

Graphics Conclusion

- A good picture is worth a 1000 words
- Encourage your analyst to provide graphical displays of the data
- These were published examples, don’t worry if your plots are simple or not as elegant
- Some journals have a graphics department that redraw graphics for a consistent look

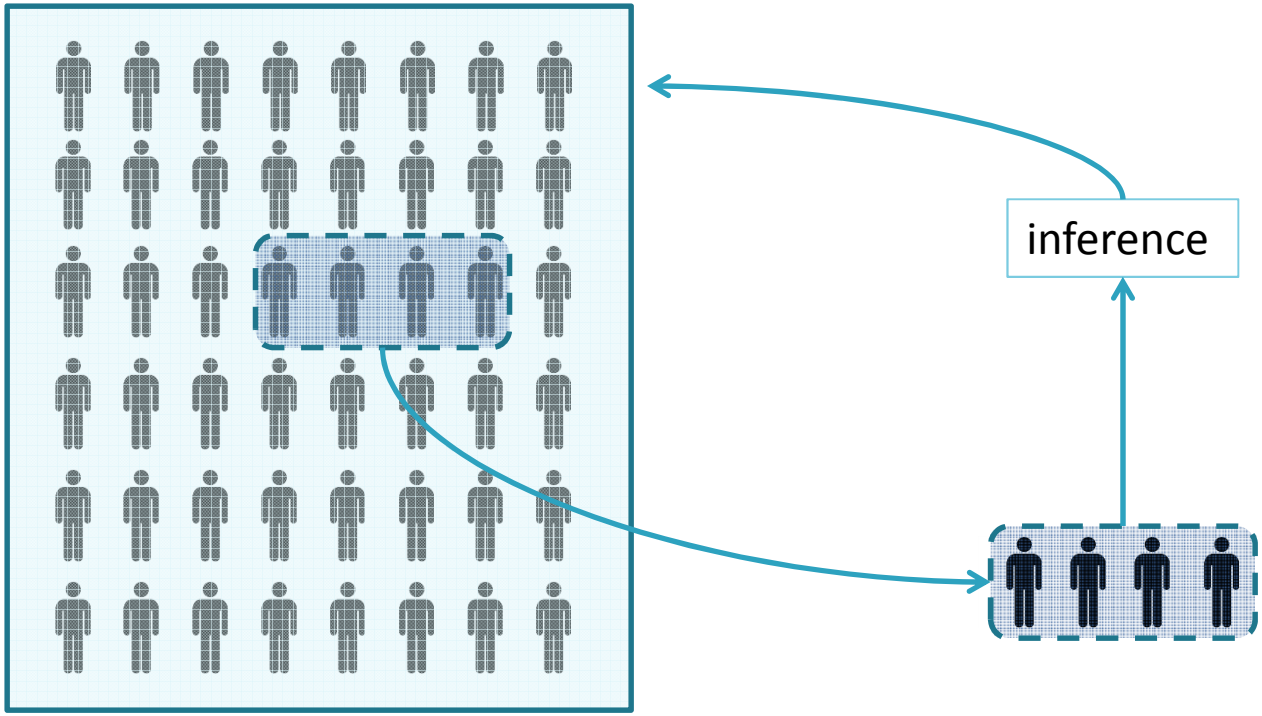
STATISTICAL INFERENCE

Confidence is what you have before you understand the problem –Woody Allen

Statistical Inference

- Drawing inference or conclusions about a population of interest,
- based on data sampled from the population,
- while appropriately accounting for uncertainty due to not having data from the entire population

Population & Sample



Statistical Inference

Many statistical methods available

Choice driven by several factors, including

- 1. Type of study being analyzed
  - Case-control study
  - Cohort study
  - Randomized clinical trial
  - Meta-analysis
  - Complex survey sample

Statistical Inference

- 1. Type of study
- 2. Type of endpoint being analyzed
  - Binary, ordinal, or continuous endpoint
  - Repeated/longitudinal measurements

Pitfall: Ignoring the independence assumption

Independent

- Single measurements from separate individuals (or animals)

Not Independent

- Repeated measure on same subject or animal
  - Before and after
  - Longitudinal
  - Assay A and B
- Measurements from separate individuals in the same household

Pitfall: Ignoring the independence assumption

- Ignoring possible dependence between observations essentially assumes the sample size is larger than it is
- Leading to anti-conservative inference, i.e., confidence intervals that are too narrow, p-values too small, and increasing the likelihood of a false-positive result

Statistical Inference

- 1. Type of study
- 2. Type of endpoint being analyzed
  - Binary, ordinal, or continuous endpoint
  - Repeated/longitudinal measurements
  - Time to event endpoint (subject to censoring) or assay (w/ limit of detection)
  - More generally, is it missing or incomplete ? Eg, due to missed visits, lost specimen, non-response, lost-to-follow-up

Pitfall: Ignoring missing data

- The default of most statistical software programs is to use only observations with complete data ("compete case" analysis)
- May be reasonable if missing data is uncommon
- Otherwise can yield invalid and/or inefficient results
- Certain types of missing data (e.g., right censoring in survival analysis) can be handled by standard software
- For other types of missing data, more sophisticated methods (imputation, weighting) may be required

Statistical Inference

- 1. Type of study
- 2. Type of endpoint being analyzed
- 3. Sample size
  - Asymptotic or exact statistical methods

Pitfall: Large sample methods with small n

- The justification for most statistical methods relies on the sample size being large
- AKA asymptotic approximations
- These approximations fail when the sample size is small (or possibly when the sample size is not small, but the endpoint is rare)
- Methods appropriate for small sample sizes (aka “exact” methods) should be used in this case

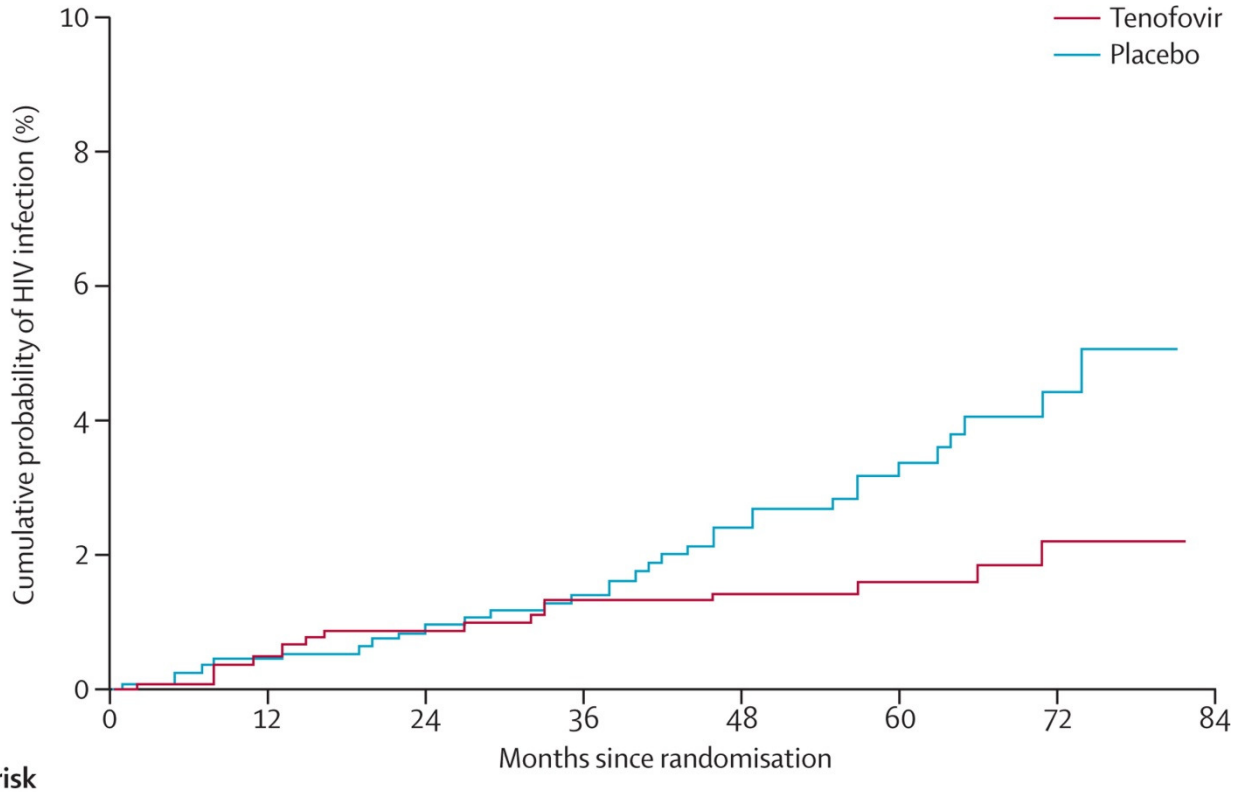
Statistical Inference

- 1. Type of study
- 2. Type of endpoint being analyzed
- 3. Sample size
- 4. Additional considerations
  - Efficiency – get the most out of the data
  - Robustness – to outliers, violations to assumptions
  - Assumptions underlying the method

Pitfall: Failure to Check Assumptions

- All statistical methods rely on certain assumptions
- If these assumptions are not true the resulting inference may be incorrect
- Thus it is critical to understand these underlying assumptions and, if possible, assess veracity for a given study/data set

Choopanya et al. Primary analysis based on Cox proportional hazards model



Number at risk							
Tenofovir	1204	1007	933	857	736	521	241
Placebo	1207	1029	948	844	722	500	234

The Lancet 2013; 381:2083-2090

What types of assumptions ?

- Parametric methods make very strong assumptions (eg log viral load is Normally distributed, time until infection follows an Exponential distribution), but are not robust
- Non-parametric methods (eg Kaplan-Meier estimator) make very few assumptions, so tend to be valid in a wide range of settings but not efficient
- Semi-parametric and rank-based type methods are popular
  - Wilcoxon ranksum test
  - Generalized estimating equations (GEE)
  - Cox proportional hazards (PH) model
- Tend to be more robust than parametric methods (i.e., make weaker assumptions) and often almost as efficient

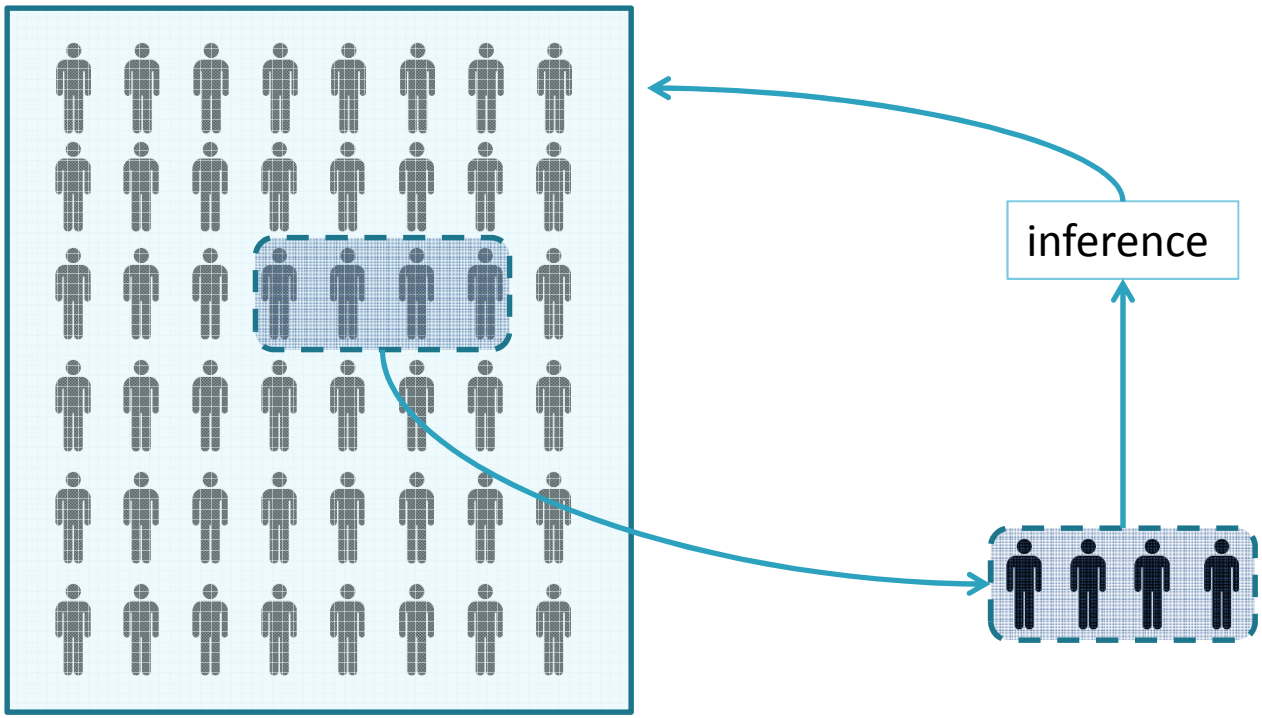
Statistical Inference

- Typically one of two forms
- Estimation: Compute an estimate and associated confidence interval
- Testing: Conduct a statistical test of a null hypothesis, p-value

Estimation

- Ideally, we'd like to know some characteristic of the population of interest (*parameter*)
- Eg, prevalence of HIV in women of child-bearing age in Malawi
- Typically we can't measure the whole population. So we sample from the population and compute an *estimate* of the parameter from the sample
- Estimate is *random* in the sense that if we were to sample from the population repeatedly, each estimate would potentially be different
- In contrast, the parameter is viewed as *fixed*

Population & Sample



Target Population
μ: mean

True population parameter

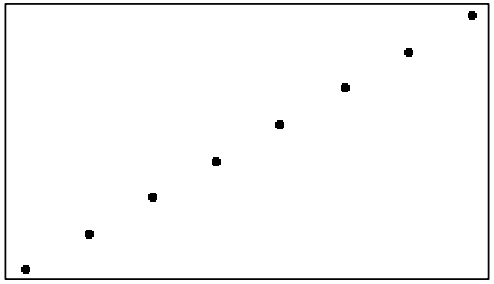
Sample
$\bar{x}$: mean

Sample estimate

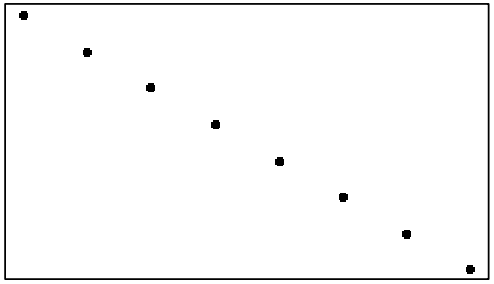
Pitfalls: Over-Reliance on Point Estimate

- Estimates alone do not describe uncertainty associated with inference; need to report confidence intervals also (more on this in a moment)
- Point estimate represents a single summary measure of the data and typically will not tell the whole story and may even mislead
- Eg, odds ratios
- Eg, next slide shows various examples where the Pearson correlation coefficient r, a measure of linear association between two variables, may or may not be misleading (know your data!)

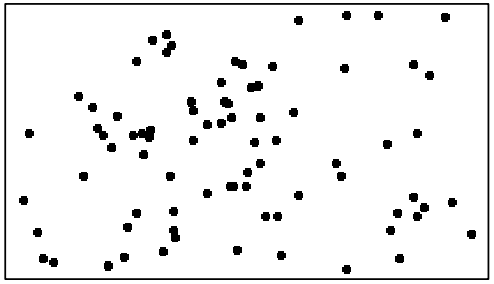
r=1



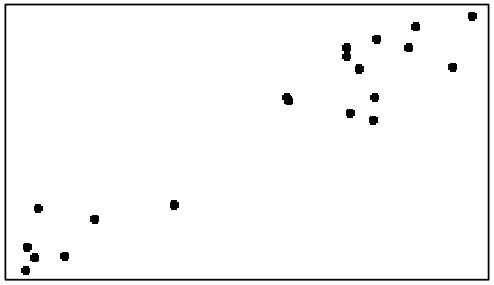
r=-1



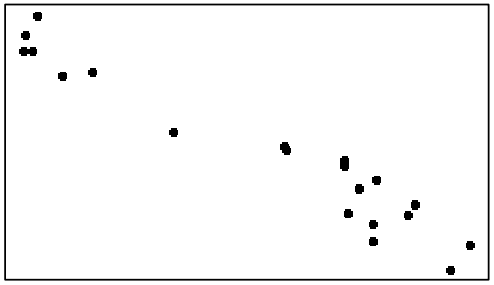
r=0



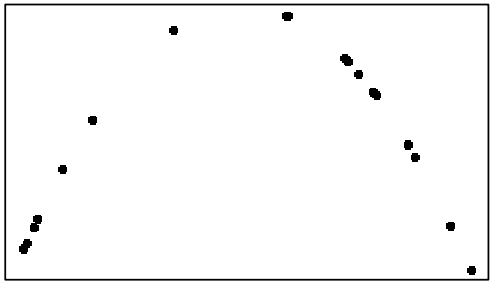
0<r<1



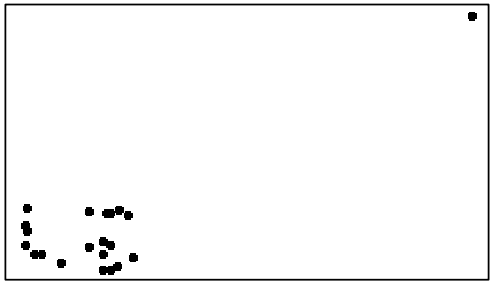
-1<r<0



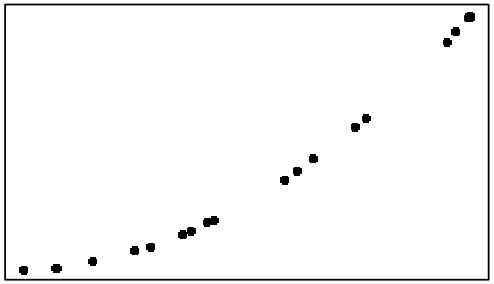
r=0



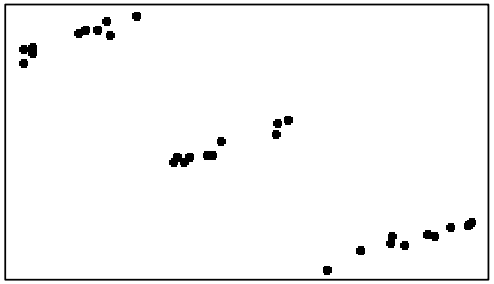
0<r<1



0<r<1



-1 < r < 0

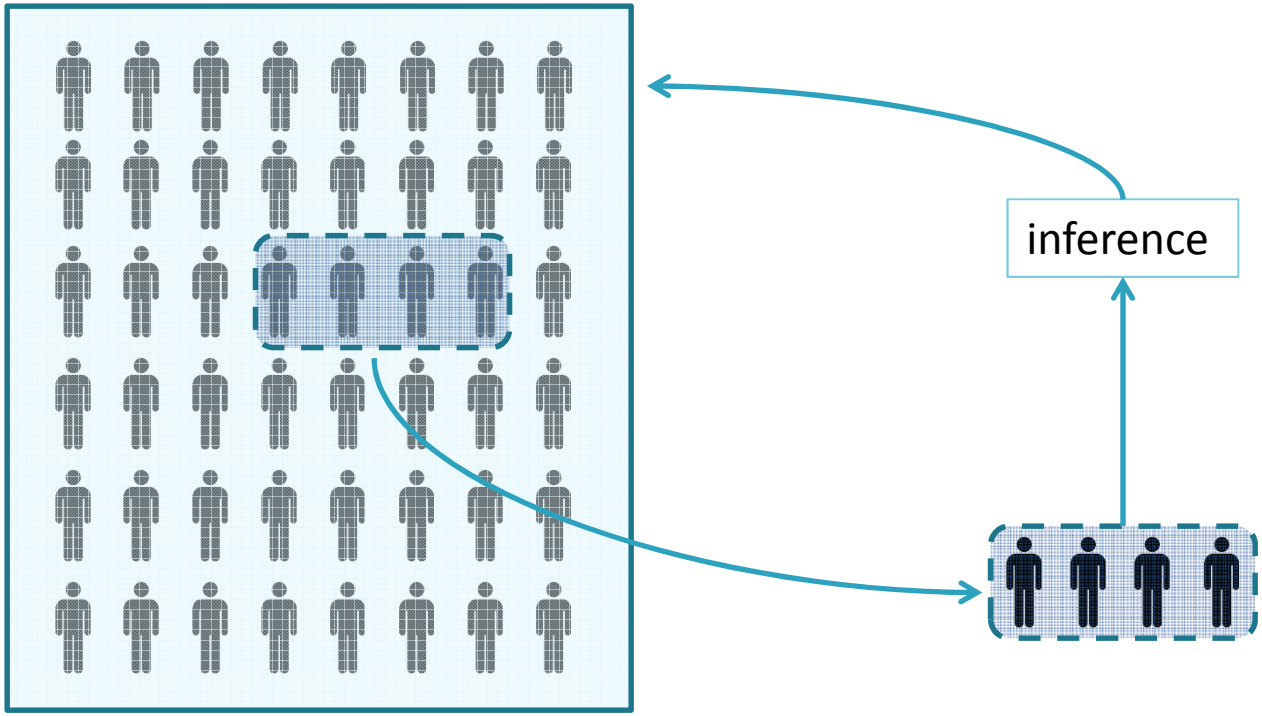


Pitfall: Confusing SD and SE

- SD: Standard Deviation
- SE: Standard Error

- Sample SD: describes the spread of the data, how much do the observations tend to deviate from the mean
- Population SD: analog of sample SD for the population; unknown; fixed; a parameter

Population & Sample



Target Population
μ: mean
σ: standard deviation

True population parameters

Sample
$\bar{x}$: mean
s: standard deviation

Sample estimates

Pitfall: Confusing SD and SE

- SD: Standard Deviation
- SE: Standard Error

- Sample SD: describes the spread of the data, how much do the observations tend to deviate from the mean
- Population SD: analog of sample SD for the population; unknown; fixed; a parameter

- SE refers to the distribution of the estimator
- Imagine computing the estimator (e.g., sample mean) repeatedly; SE is the standard deviation of the different estimators

What is a confidence interval (CI)?

- Provides an indication of the uncertainty associated with an observed estimate
- How to interpret ? 95% CI: Imagine conducting 100 studies, 95 of the 100 intervals should contain the true parameter of interest
- To achieve narrower confidence intervals, decrease confidence (90% CI will be narrower than a 95% CI) or increase sample size

Reporting Results

- General recommendation is to report estimate as well as confidence interval

- Cohen et al (2011) “a total of 39 HIV-1 transmissions were observed (incidence rate, 1.2 per 100 person-years; 95% confidence interval [CI], 0.9 to 1.7)”

Statistical Inference

- Typically one of two forms
- Estimation: Compute an estimate and associated confidence interval
- Testing: Conduct a statistical test of a null hypothesis, p-value

Hypothesis Testing

- Stipulate a null hypothesis
- Construct a test statistic based on data from sample, typically defined in such a way to help distinguish whether or not the null might be true
- Assess likelihood of obtaining a statistic as or more extreme than the observed statistic under the null: p-value
- If the p-value is small (e.g., <0.05), one of two possibilities:
  - Just witnessed something unusual
  - The null hypothesis is not true
- Reject the null if likelihood is v small (say <.05)
- Threshold for rejecting null: significance level α

Example: BAN Study (Chasela et al NEJM 2010)

- Evaluate efficacy of a maternal triple-drug antiretroviral regimen or infant nevirapine prophylaxis for 28 weeks during breast-feeding to reduce postnatal HIV transmission in Malawi
- Two primary null hypotheses of interest:

  (1) the cumulative risk of HIV by 28 weeks will be the same in the infant intervention arm as in the no-ARV arm

  (2) the cumulative risk of HIV by 28 weeks will be the same in the maternal intervention arm as in the no-ARV arm

- Log-rank test used to test each null hypothesis
- $P<0.001$ and $P=0.02$; reject both nulls

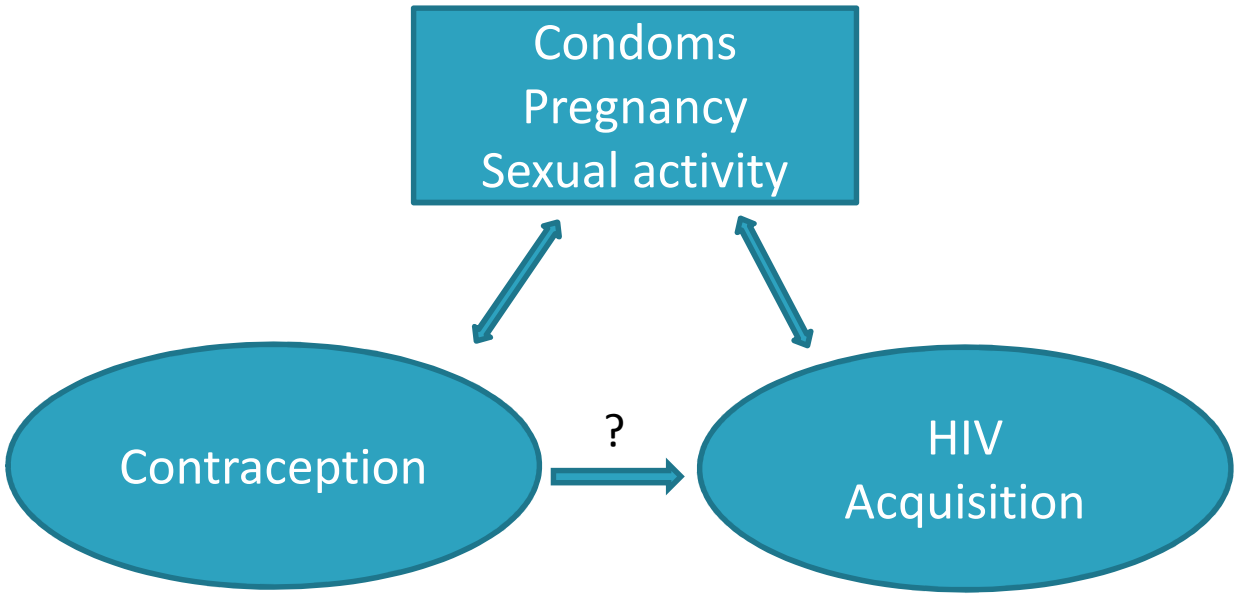
Confidence interval (CI) and p-value

- Generally recommended to not report p-values alone; reduces entire data set to a single number
- Accompany with parameter estimate and CI
- Cohen et al (NEJM 2011) “Of the 28 linked transmissions, only 1 occurred in the early-therapy group (hazard ratio, 0.04; 95% CI, 0.01 to 0.27; P<0.001)”
- Jamieson et al (Lancet 2012) “The cumulative risk of HIV-1 transmission by 48 weeks was significantly higher in the control group (7%, 95% CI 5–9) than in the maternal-antiretroviral (4%, 3–6; p=0·0273) or the infant-nevirapine (4%, 2–5; p=0·0027) groups.”

Pitfall: Over-emphasis on “significance”

- P<0.05 usually accompanied by phrase “statistically significant”
- Large effects in a small sample may be scientifically important but not statistically significant (underpowered)
- Small effects in a large sample can be statistically significant (high power to detect small effects), but **not** scientifically/clinically meaningful

Pitfall: Significant association implies causality

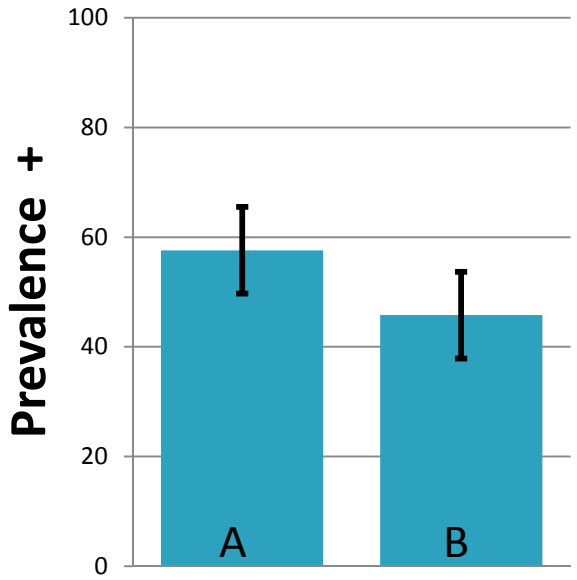


STATISTICAL INFERENCE

A Simple Example

Comparing Proportions

Frequency Table			
	Group		
Outcome	A	B	Total
+	87	70	157
--	64	83	147
Total	151	153	304



Estimated Prevalence of "+"
(95% confidence interval)

A: 57.6% (49.7 to 65.5%)
B: 45.8% (37.9 to 53.7%)

Prevalence is 11.8% higher in group A.

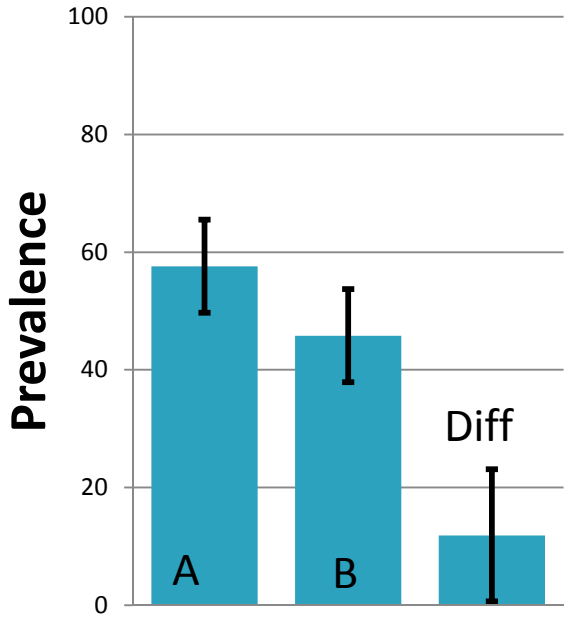
Is the difference statistically significant?

Pitfall: Confidence Interval Overlap

- For the two sample problem, CIs overlapping does not indicate a lack of significant difference
- Need to conduct a hypothesis test or compute CI for difference

Comparing Proportions

Frequency Table			
	Group		
Outcome	A	B	Total
+	87	70	157
--	64	83	147
Total	151	153	304



Group A: 57.6%
Group B: 45.8%

The estimated difference is 11.8% with a 95% confidence interval from 0.7% to 23.0%

Chi-square p=0.04

At the 0.05 level of significance, there is a significant difference in prevalence between the two groups.

STATISTICAL INFERENCE

Multiple Comparisons

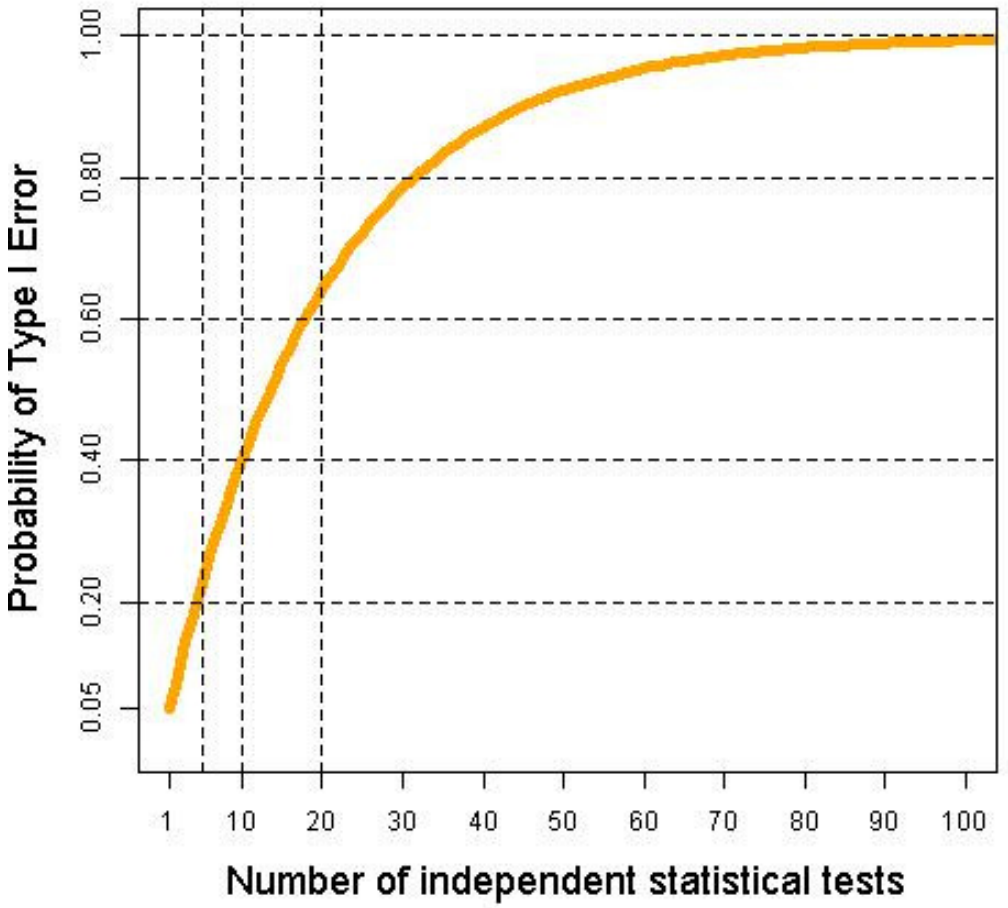
Multiple Comparisons

- Suppose we conduct many hypothesis tests, using data from the same study or individual
- Eg, compare group A versus group B with respect to several different assays, several different output measurements from the same assay (eg, flow cytometry), large panel of lab tests
- What is the probability of making at least one type I error (false positive) ?
- Eg, suppose we conduct 10 tests, each with $\alpha=0.05$, and in truth all nulls are true. What is the chance of making at least one type I error (falsely rejecting the null) ?

Multiple Comparisons

Familywise Error Rate: probability of getting a significant result ($p<0.05$) by chance when there are multiple independent tests and no correction for the number of tests carried out

10 independent tests:
40% probability
of a type I error
(false positive)



Reference: Intuitive Biostatistics

Multiple Comparisons

- Methods are available to adjust for MC
  - Bonferroni adjustment
    - divide α by the number of tests to be performed
    - appropriate for independent tests, conservative otherwise
  - Hochberg, Holm, …
  - False Discovery Rate (FDR) adjustment
    - Less stringent, useful when a lot of tests are done (genomics)
- When to adjust for MC ?
- In settings such as genomics where a large number of tests are performed, MC adjustment essential

Multiple Comparisons

- In randomized studies with more than two arms, MC adjustment standard
  - BAN: A log-rank test procedure of size 0.025 used to test each null hypothesis
- Analyzing the same outcome multiple ways to assess the sensitivity of the primary approach can be done without MC adjustment
- Conducting multiple tests without MC adjustment and reporting only “significant” results not appropriate
- Minimal approach: report all test conducted and state the results are “without correction for multiple comparisons” and make clear how many comparisons were carried out. Informally, the reader can expect 1 in 20 $p<0.05$ to be from chance

STATISTICAL INFERENCE

Choosing an Endpoint

Choosing an Endpoint

- To avoid post-hoc data dredging, important to select primary endpoint(s) prior to analysis
- Considerations when selecting an endpoint:
  - Relevance
    - What is the most important question to answer ?
  - Reliable (precise) and valid (accurate)
    - Substantial measurement error or bias ?
  - Rate
    - How often does the endpoint occur (e.g., incident HIV infection) in the population of interest ? This will affect power (more to come)

Types of Endpoints

- Clinical endpoints
  - HIV infection, AIDS-defining event, death
  - Eg, HIV vaccine efficacy studies use HIV infection as endpoint
- Surrogate endpoints
  - Not always feasible or ethical to wait for clinical endpoints
  - Laboratory: HIV-1 RNA, CD4, lipids, biomarkers…
  - Eg, early phase HIV vaccine studies rely on immunogenicity endpoints
  - Potentially misleading, ie, effect on surrogate may not imply effect on clinical endpoint
- Composite endpoints
  - Eg, time until HIV infection or death (BAN)
  - More relevant from public health or policy standpoint ?
  - Can simplify analysis and increase power, but results may be difficult to interpret

SAMPLE SIZE AND POWER

Sample Size and Power Calculation

- Most grants/protocols include sample size justification

- Sample size too small: may miss scientifically meaningful differences
- Sample size too big: waste of resources

- Sample size justification typically in terms of power

Type I and Type II Error: Hypothesis Testing

Population (Truth)	DECISION	
Null hypothesis…	Reject the null	Do not reject the null
Is True	Type I error (α) "false positive"	Correct
Is False	Correct "power"	Type II error (β) "failure to detect" "false negative"

Typically in study design α set to 5% and power (1-β) set to ≥80%

Sample Size/Power Calculation

- Based on endpoints associated with primary aims
- Requires prior knowledge (or a guess) regarding the endpoints (eg, mean, rate, SD) in each arm of the study
- Power is a function of exact specification of the alternative hypothesis; sometimes called the "effect size"
- Effect size chosen based on scientific or clinical relevance, not by the statistician

Sample Size/Power Calculation

- Takes one of two forms:
  1. For fixed sample size, what is the power to detect a particular difference between the arms of the study (ie effect size) ?
  2. Given a particular difference (effect size) that we would like to detect, what sample size is required to insure adequate power ?

Generally three approaches

- 1. Mathematical formula in a book/article
- 2. Statistical software (which often relies on 1)
- 3. Simulation study (modeling)

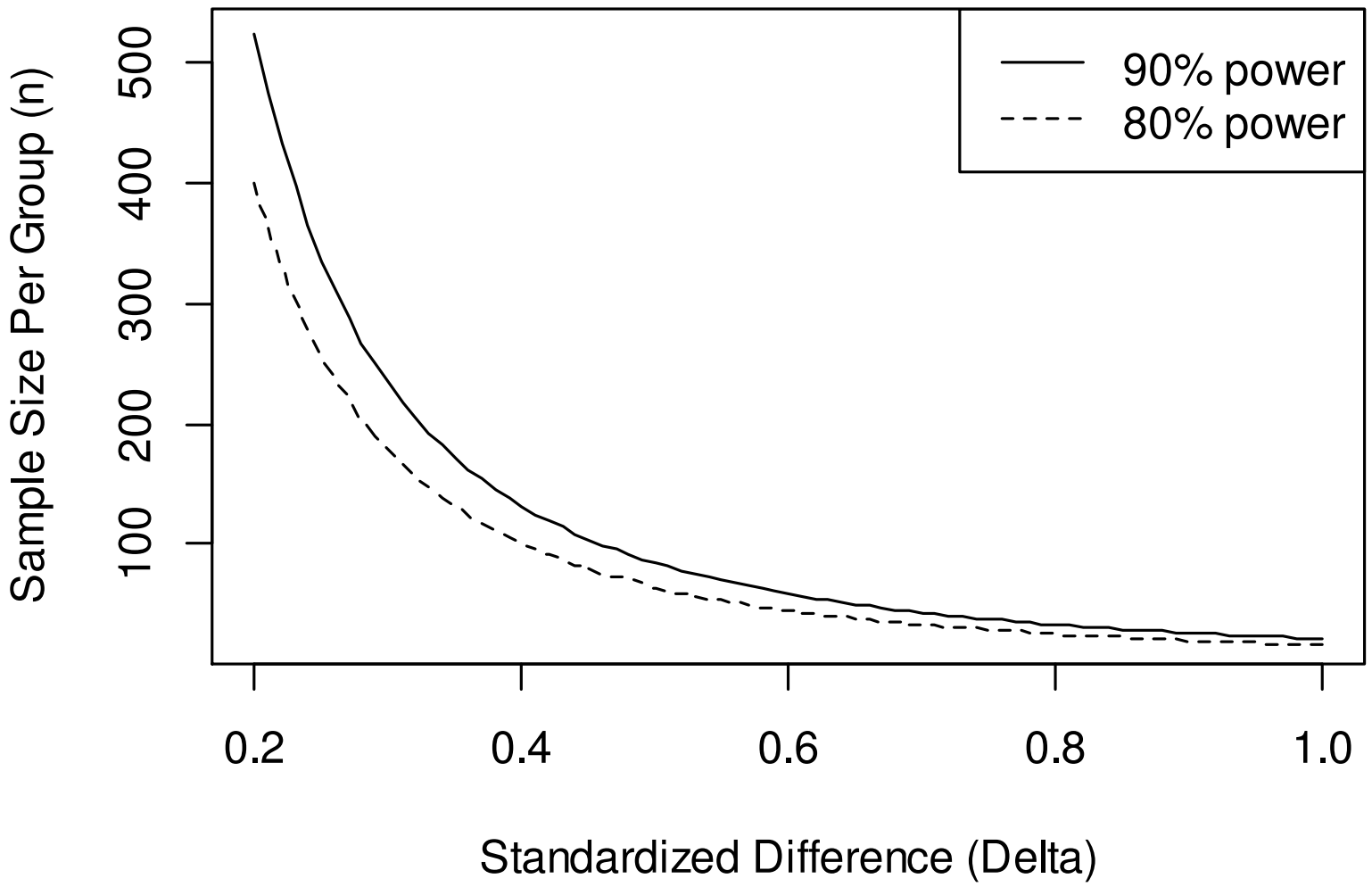
Sample Size Formula: Comparing Two Means

- Continuous outcome, two groups, alpha = 0.05, power = 80%
- Sample size per group

$$n=16/\Delta^2$$

  where $\Delta=(\mu_0-\mu_1)/\sigma$ is the difference in means between the two groups (effect size) divided by the population SD, i.e., standardized difference
- Note for continuous outcome need to know SD in addition to effect size
- For 90% power, replace 16 with 21

Gerald van Belle. Statistical Rules of Thumb. Chapter 2


90% power
80% power
Sample Size Per Group (n)
100
200
300
400
500
0.2
0.4
0.6
0.8
1.0
Standardized Difference (Delta)

What impacts n?

- Effect size of interest or # of events
  - The smaller the effect size (# of events), the larger n will need to be
- Variability of outcome measure
  - The greater the variability, the larger the n will need to be
- Alpha and power (e.g. 5% and 80%)
  - The smaller alpha or the greater the power, the larger n will need to be
- Number of groups and multiple comparisons
  - The greater the number of MC, the larger n will need to be

RESOURCES & REFERENCES

Statistical Software

- SAS: industry standard, legacy, learning curve
  - Useful for data management and statistics
- R: **free**, flexible, cutting edge methods, learning curve, http://www.r-project.org/
  - R Packages: akin to the peer-reviewed Wiki world of statistics
  - R Commander, Graphical User Interface (GUI) http://www.rcommander.com/
  - S-plus: proprietary version, very similar code
- STATA: GUI or syntax, more tailored to epidemiology and biostatistics

Statistical Software

- GraphPad Prism: appears user-friendly and has some built in statistical guidance
  - QuickCalcs: **free** quick analysis on the fly
  - http://www.graphpad.com/quickcalcs/
  - "Intuitive Biostatistics" –Harvey Motulsky
- Many, many more: StatXact, SigmaPlot, SPSS, Minitab, SAS JMP…
- Microsoft Excel: spreadsheet software
  - Quick graphs (use chart layout tools carefully for best results)
  - Reasonable for simple graphics and analyses

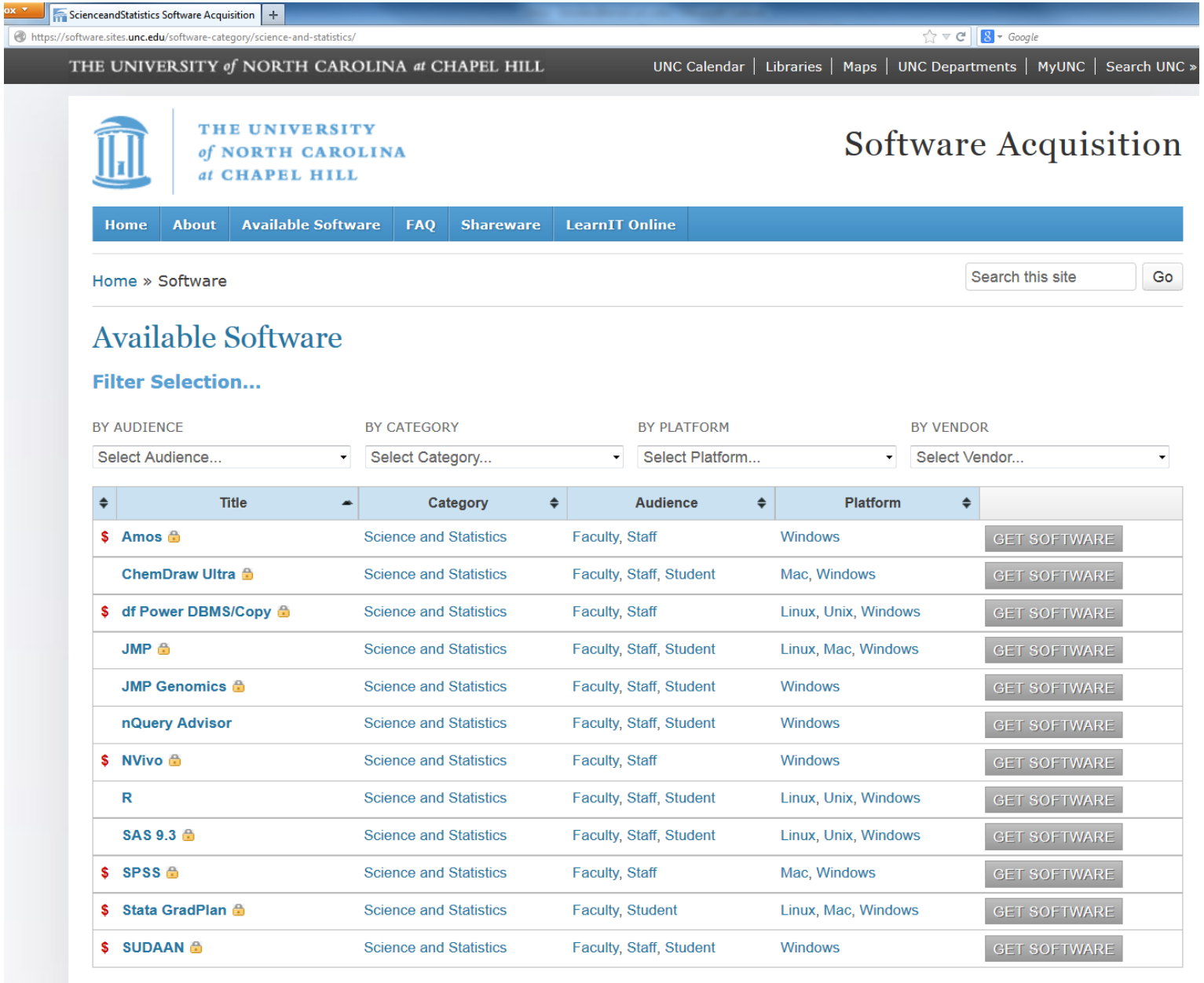
Statistical Software

- For sample size/power
  - nQuery software available for free at UNC
  - SAS PROC POWER

- Software available for free or discount at UNC

  http://software.sites.unc.edu/software-category/science-and-statistics/

  - **SAS, SAS JMP**, and R (R is freeware)
  - Discounts for several software packages (Stata, SPSS,..)
  - **Sample size calculation: nQuery Advisor**


ScienceandStatistics Software Acquisition
https://software.sites.unc.edu/software-category/science-and-statistics/
THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL
UNC Calendar | Libraries | Maps | UNC Departments | MyUNC | Search UNC »
THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL
Software Acquisition
Home | About | Available Software | FAQ | Shareware | LearnIT Online
Home » Software
Search this site
Go
Available Software
Filter Selection...
BY AUDIENCE
Select Audience...
BY CATEGORY
Select Category...
BY PLATFORM
Select Platform...
BY VENDOR
Select Vendor...
Title
Category
Audience
Platform
$ Amos
Science and Statistics
Faculty, Staff
Windows
GET SOFTWARE
ChemDraw Ultra
Science and Statistics
Faculty, Staff, Student
Mac, Windows
GET SOFTWARE
$ df Power DBMS/Copy
Science and Statistics
Faculty, Staff
Linux, Unix, Windows
GET SOFTWARE
JMP
Science and Statistics
Faculty, Staff, Student
Linux, Mac, Windows
GET SOFTWARE
JMP Genomics
Science and Statistics
Faculty, Staff, Student
Windows
GET SOFTWARE
nQuery Advisor
Science and Statistics
Faculty, Staff, Student
Windows
GET SOFTWARE
$ NVivo
Science and Statistics
Faculty, Staff
Windows
GET SOFTWARE
R
Science and Statistics
Faculty, Staff, Student
Linux, Unix, Windows
GET SOFTWARE
SAS 9.3
Science and Statistics
Faculty, Staff, Student
Linux, Unix, Windows
GET SOFTWARE
$ SPSS
Science and Statistics
Faculty, Staff
Mac, Windows
GET SOFTWARE
$ Stata GradPlan
Science and Statistics
Faculty, Student
Linux, Mac, Windows
GET SOFTWARE
$ SUDAAN
Science and Statistics
Faculty, Staff, Student
Windows
GET SOFTWARE

References and Resources

- Intuitive Biostatistics –Motulsky
- Principles of Biostatistics –Pagano
- Modern Epidemiology –Rothman
- Statistics notes in BMJ

http://www-users.york.ac.uk/~mb55/pubs/pbstnote.htm

- Bios 662

http://www.bios.unc.edu/~mhudgens/bios/662/2008fall/bios662.html

- CFAR Biostatistics Core (including these slides)

http://cfar.med.unc.edu/content/biostatistics-core

Biostatistics Training at UNC

Odum Institute
http://www.irss.unc.edu/odum/home2.jsp

Biostatistics Department
http://sph.unc.edu/department-pages/biostatistics/

BIOS 600 PRINCIPLES OF STATISTICAL INFERENCE (3). Prerequisite, knowledge of basic descriptive statistics. Major topics include elementary probability theory, probability distributions, estimation, tests of hypotheses, chi-squared procedures, regression, and correlation.

BIOS 610 BIOSTATISTICS FOR LABORATORY SCIENTISTS (3). Prerequisite, elementary calculus. Introduces the basic concepts and methods of statistics, focusing on applications in the experimental biological sciences.

NC TraCS Biostatistics Training

Annually: Introduction to Study Design and Strategies for Data Analysis (1 week summer course)

This Fall: Biostatistics Seminar Series for Clinical and Translational Scientists begins August 29!

A seven part series running Aug-Nov

tracs.unc.edu/biostats-seminar-fall2013

NC TraCS Biostatistics Training

The **goal** of this series is to provide clinical and translational researchers who have basic quantitative training in biostatistical methods with a more in depth understanding of selected topics and to introduce them to more advanced methods.

The **target audience** for the seminar series includes junior clinical and translational scientists who currently serve as or plan to serve as Principal Investigators leading interdisciplinary research teams that include biostatisticians. For example, this includes junior faculty in the Schools of Medicine, Public Health, Nursing, Pharmacy and Dentistry.

Fall 2013 NC TracS Biostatistics Seminars

To receive information about future sessions, contact NCTraCS_BiosSeminar@unc.edu.

2013 - 2014 Sessions

Date / Time	Topic / Location	Presenter
Thu. 8/29/13 1:00 - 2:30	Some Essentials of Randomized Controlled Trials Brinkhous-Bullitt 219	Mark Weaver, PhD UNC School of Medicine
Wed. 9/12/13 1:00 - 2:30	Data Reliability and Validity Brinkhous-Bullitt 219	Kant Bangdiwala, PhD UNC Gillings School of Global Public Health
Thu. 9/26/13 1:00 - 2:30	DSMBs and Interim Analyses Bondurant Hall, G074	Sonia Davis, DrPH UNC Gillings School of Global Public Health, CSCC

http://tracs.unc.edu/events/biostatistics-seminar-series-for-clinical-and-translational-scientists.html

Fall 2013 NC TracS Biostatistics Seminars

Thu. 10/10/13 1:00 - 2:30	Cluster-Randomized Trials Brinkhous-Bullitt 219	John Preisser, PhD UNC Gillings School of Global Public Health
Thu. 10/24/13 1:00 - 2:30	Issues in Non-inferiority Trials Bondurant Hall, G074	Rosalie Dominik, DrPH UNC Gillings School of Global Public Health, CSCC
Thu. 11/7/13 1:00 - 2:30	Survey Sampling and Working with the CSRL Brinkhous-Bullitt 219	Donglin Zeng, PhD, and Robert Agans, PhD Co-Directors of Carolina Survey Research Laboratory (CSRL)
Thu. 11/21/13 1:00 - 2:30	Design and Analysis of SMARTs (Sequential Multiple Assignment Randomized Trials) Brinkhous-Bullitt 219	Michael Kosorok, PhD UNC Gillings School of Global Public Health

Thank you for coming and staying !

Thanks to Joe Rigdon, Ali Fokar, David Rosen, Catherine Grodensky, Carol Golin for helpful feedback !

Next Week:

"An Overview of Data Management"

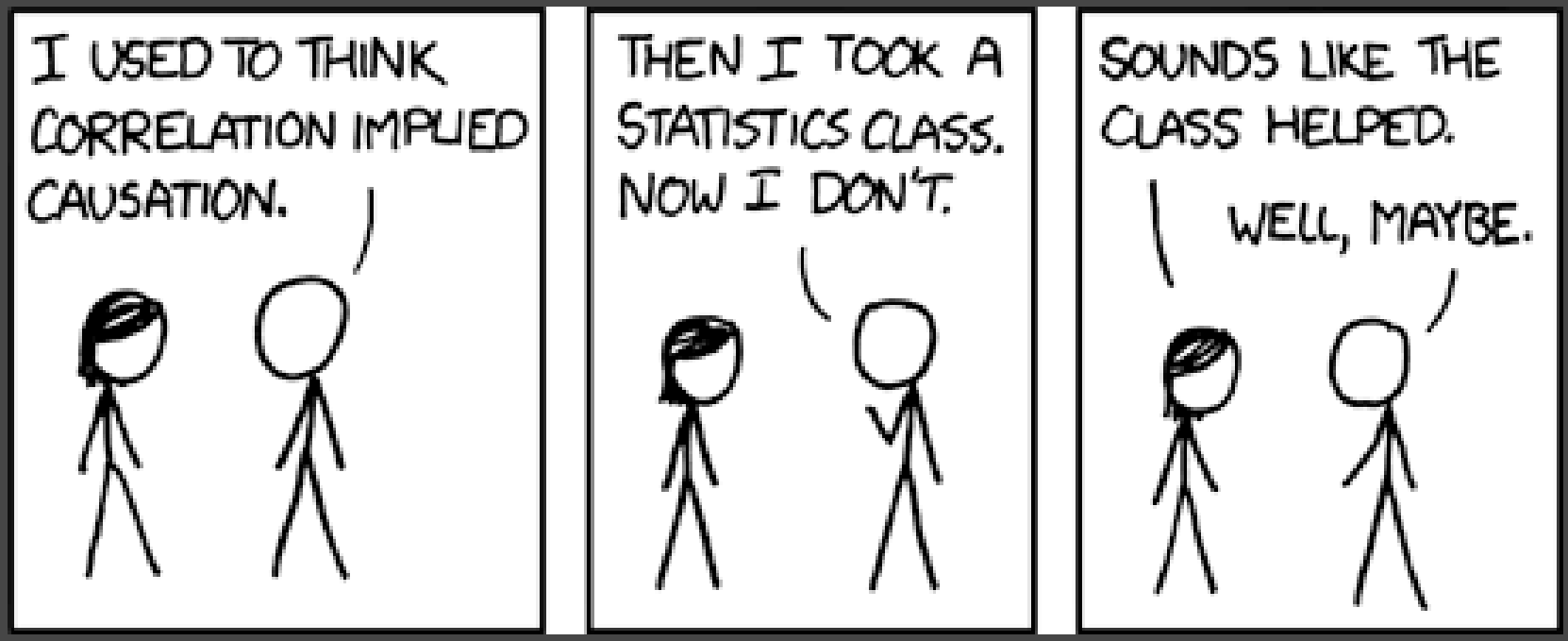
Ali Foker
UNC CFAR Clinical Core

and

"Survey Development"

Carol Golin and Catherine Grodensky
UNC CFAR Social and Behavioral Science Core

August 30, 2013
8:30 - 10:00 a.m. (1.5 hours)
1131 Bioinformatics


I USED TO THINK CORRELATION IMPLIED CAUSATION.
THEN I TOOK A STATISTICS CLASS. NOW I DON'T.
SOUNDS LIKE THE CLASS HELPED.
WELL, MAYBE.