



Stats Overview for Clinical Researchers

Frank E Harrell Jr

VICTR Design Biostatistics and Research Ethics Program
Department of Biostatistics
Vanderbilt University School of Medicine

22 April 2011



Goal

To obtain unambiguous results at minimum cost

- Perform experiment to provide a valid estimate of the underlying variability of responses
- Use randomization to obtain estimates of variability and balance on uncontrolled variables
- Use blocking to balance out known sources of variation not of primary interest
- Use factorial designs to learn about the effects of multiple factors in a single more complex experiment, rather than a separate experiment (in different subjects or animals) for each experimental manipulation



Refining Measurements

Stats
Overview for
Clinical
Researchers

Design

Descriptive
Stats

Hypothesis
Testing

Bias and
Precision

Presenting
Results

Bayesian
Methods

Respecting
Continuous
Variables

- Optimize clinical/physiologic/anatomical meaning
- Optimize information content
 - Continuous > ordinal > binary response
- Minimize technical and observer variability
- Minimize damage caused by missing data, LLD
- Measure covariates efficiently and adjust for them



- Best not to assume shape of distributions
- Let the data speak for themselves
- Three-number summary: 25th, 50th (median), 75th percentiles
- Describes central tendency, spread, symmetry



Hypothesis Testing

Stats
Overview for
Clinical
Researchers

Design

Descriptive
Stats

Hypothesis
Testing

Bias and
Precision

Presenting
Results

Bayesian
Methods

Respecting
Continuous
Variables

- Existence of ESP is a hypothesis
- Assessing effects of drugs, procedures, devices involves estimation
- Many studies powered to detect huge effect
- If effect is not huge, no information from study



- Provide evidence against a *null* hypothesis
- Probability of a statistic as impressive as yours **if** H_0 true
- Not a probability of an effect or difference (same problem with sensitivity)
- **No** conclusion possible from large P -values
- Cannot conclude clinical relevance from small P



- Best addressed with study design
 - randomization
 - minimize work-up or referral bias
- Sometimes handled by careful regression analysis
 - adjust for patient selection
 - adjust for confounding risk factors



- Erroneous estimates caused by bias and imprecision
- Precision = margin of error
- Standard error or $\frac{1}{2}$ width of confidence interval if estimate is unbiased
- Margin of error \downarrow as $n \uparrow$



How Not to Present Results

- $P = 0.02$ — let's put this into clinical practice
- $P = 0.4$ — this drug does not kill people
- $P = 0.2$ but there is a trend in favor of our blockbuster drug
- The observed difference was 6mmHg and we rejected H_0 so the true effect is 6mmHg.



How Not to Present Results, cont.

Stats
Overview for
Clinical
Researchers

Design

Descriptive
Stats

Hypothesis
Testing

Bias and
Precision

Presenting
Results

Bayesian
Methods

Respecting
Continuous
Variables

- The proportion of patients having adverse events was 0.01 and 0.03; the study wasn't powered to detect adverse event differences so we present no statistical analysis
- The reduction in blood pressure was 6mmHg with 0.95 C.L. of [1mmHg, 11mmHg]; the drug is just as likely to only reduce blood pressure by 1mmHg as it is by 6mmHg.



How to Present Results

Stats
Overview for
Clinical
Researchers

Design

Descriptive
Stats

Hypothesis
Testing

Bias and
Precision

Presenting
Results

Bayesian
Methods

Respecting
Continuous
Variables

- Estimates should be accompanied by confidence limits
- Confidence limits can be computed without regard to sample size or power
- A computed value from a sample is only an estimate of the population value
- Best to think of an estimate from a study as a fuzz, not a point



Stats
Overview for
Clinical
Researchers

Design

Descriptive
Stats

Hypothesis
Testing

Bias and
Precision

Presenting
Results

Bayesian
Methods

Respecting
Continuous
Variables



Copyright Jim Fink

www.phinxphotography.com/paintbrush_dandelion.htm



Bayesian Presentation

Stats
Overview for
Clinical
Researchers

Design

Descriptive
Stats

Hypothesis
Testing

Bias and
Precision

Presenting
Results

Bayesian
Methods

Respecting
Continuous
Variables

- Posterior probability density (like histogram) for likelihood of effects equaling certain values
- Solves “optical illusion” problem of flat confidence bars
- Is the most intuitive way to communicate evidence
- Can compute probability of a *clinically significant* difference



Bayesian Current Probability Distribution

Stats
Overview for
Clinical
Researchers

Design

Descriptive
Stats

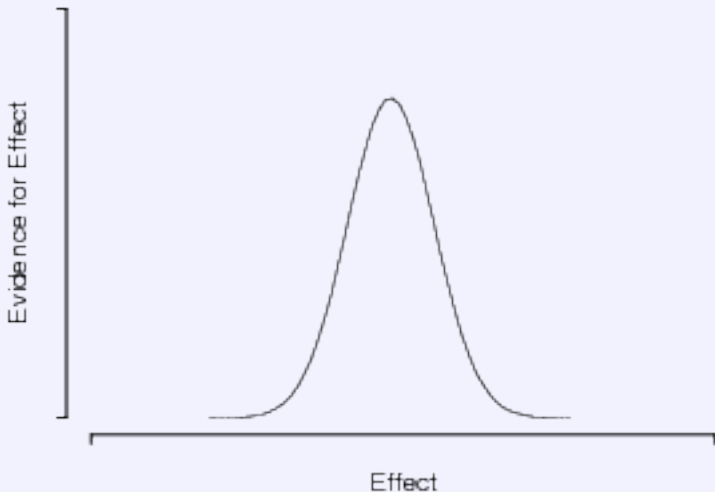
Hypothesis
Testing

Bias and
Precision

Presenting
Results

Bayesian
Methods

Respecting
Continuous
Variables





Example: Comparing Two Proportions

Stats
Overview for
Clinical
Researchers

Design

Descriptive
Stats

Hypothesis
Testing

Bias and
Precision

Presenting
Results

Bayesian
Methods

Respecting
Continuous
Variables

- Provide the two proportions
- Confidence limits for difference
- Confidence limits for relative difference (odds ratio)
- Bayesian posterior probabilities of these two



Respecting Continuous Variables

Stats
Overview for
Clinical
Researchers

Design

Descriptive
Stats

Hypothesis
Testing

Bias and
Precision

Presenting
Results

Bayesian
Methods

Respecting
Continuous
Variables

- Keep all continuous variables continuous
- Maximizes power and precision
- Cut-points are arbitrary
- Diagnosis: use extent of disease instead of presence
- Prognosis: days until clinical endpoint
- Test output: use actual measurements or degree of positivity



Joint Effect of Age and Cholesterol on Risk of CAD

Stats
Overview for
Clinical
Researchers

Design

Descriptive
Stats

Hypothesis
Testing

Bias and
Precision

Presenting
Results

Bayesian
Methods

Respecting
Continuous
Variables

