RCTdesign

Biost / Stat 579: Data Analysis and Report Writing

Scott S. Emerson, M.D., Ph.D.
Professor of Biostatistics
University of Washington

Lecture 1: Preliminaries

September 30, 2011

1

RCTdesign

Preliminaries

Course Organization

2

Course Objectives

- Learn an organized approach to the analysis of data
- Gain experience in the analysis of data to answer scientific questions
- Gain experience in writing reports of analyses to statistically naïve collaborators

3

Course Format

- Homeworks directed to selected portions of a total analysis
  - Statistical Analysis Plans
  - Analysis and reports
- Course discussion
  - Presentation and critique of SAPs and analyses/reports
- Final project
  - An analysis and report in the format of the PhD qualifying exam
  - Oral defense

4

Resources

- Course webpage: www.emersonstatistics.com/b579/
- Approach to analyzing a data set
- Homework assignments, data sets

5

Major Emphasis

- Thinking about the problem
- Statistical Analysis Plan
- Statistical analyses having minimal assumptions
- Clear and straightforward description of analyses

6

Statistical Analysis Plan

- A complete pre-specification of all analyses that will be performed to answer the scientific question of interest
  - Sampling scheme
  - Description of available data
  - Analysis populations
  - Analysis models
    - Outcome variables
    - Summary measure
    - Link function
    - Exact form of covariates
    - Exact form of statistics
  - Tables and figures

7

Clinical Trials

- Experimentation in human volunteers
- Investigates a new treatment/preventive agent
  - Safety:
    - Are there adverse effects that clearly outweigh any potential benefit?
  - Efficacy:
    - Can the treatment alter the disease process in a beneficial way?
  - Effectiveness:
    - Would adoption of the treatment as a standard affect morbidity / mortality in the population?

8

U.S. Regulation of Drugs / Biologics

- Wiley Act (1906)
  - Labeling
- Food, Drug, and Cosmetics Act of 1938
  - Safety
- Kefauver – Harris Amendment (1962)
  - Efficacy / effectiveness
    - " [If] there is a lack of substantial evidence that the drug will have the effect ... shall issue an order refusing to approve the application. "
    - "...The term 'substantial evidence' means evidence consisting of **adequate and well-controlled investigations, including clinical investigations**, by experts qualified by scientific training"
- FDA Amendments Act (2007)
  - Registration of RCTs, Pediatrics, Risk Evaluation and Mitigation Strategies (REMS)

9

Medical Devices

- Medical Devices Regulation Act of 1976
  - Class I: General controls for lowest risk
  - Class II: Special controls for medium risk - 510(k)
  - Class III: Pre marketing approval (PMA) for highest risk
    - "…**valid scientific evidence** for the purpose of determining the safety or effectiveness of a particular device … adequate to support a determination that there is reasonable assurance that the device is safe and effective for its conditions of use…"
    - "Valid scientific evidence is evidence from well-controlled investigations, partially controlled studies, studies and objective trials without matched controls, well-documented case histories conducted by qualified experts, and reports of significant human experience with a marketed device, **from which it can fairly and responsibly be concluded by qualified experts that there is reasonable assurance of the safety and effectiveness**…"
- Safe Medical Devices Act of 1990
  - Tightened requirements for Class 3 devices

10

The Problem of Clinical Trial Design

Happy families are all alike; every unhappy family is unhappy in its own way.

Leo Tolstoy, *Anna Karenina, 1873-77*

11

The Problem of Clinical Trial Design

Unbiased clinical trials are all alike; every **biased clinical trial** is **biased** in its own way.

12

Clinical Trial Design

- Finding an approach that best addresses the often competing goals: Science, Ethics, Efficiency
  - Basic scientists: focus on mechanisms
  - Clinical scientists: focus on overall patient health
  - Ethical: focus on patients on trial, future patients
  - Economic: focus on profits and/or costs
  - Governmental: focus on safety of public: treatment safety, efficacy, marketing claims
  - Statistical: focus on questions answered precisely
  - Operational: focus on feasibility of mounting trial

13

Statistical Planning

- Satisfy collaborators as much as possible
- Discriminate between relevant scientific hypotheses
  - Scientific and statistical credibility
- Protect economic interests of sponsor
  - Efficient designs
  - Economically important estimates
- Protect interests of patients on trial
  - Stop if unsafe or unethical
  - Stop when credible decision can be made
- Promote rapid discovery of new beneficial treatments

14

Classical Fixed Sample Designs

- Design stage:
  - Choose a sample size
- Conduct stage:
  - Recruit subjects, gather all the data
- Analysis stage:
  - When all data available, analyze and report

15

Statistics and Science

- Statistics is about science
  - Science in the broadest sense of the word
- Science is about proving things to people
  - Science is necessarily adversarial
    - Competing hypotheses to explain the real world
  - Proof relies on willingness of the audience to believe it
  - Science is a process of successive studies
- Game theory: Accounting for conflicts of interest
  - Financial
  - Academic / scientific

16

Science vs Statistics

- Recognizing the difference between
  - The parameter space
    - What is the true scientific relationship?
  - The sample space
    - What data will you / did you gather?

17

"Parameter" vs "Sample" Relationships

- The true scientific relationship ("parameter space")
  - Summary measures of the effect in population
    - Means, medians, geometric means, proportions…
- Scientific "sample space" scales:
  - Estimates attempting to assess scientific importance
    - Point estimate is a statistic estimating a "parameter"
    - Interval estimates
      - CI describes the values in the "parameter space" that are consistent with the data observed (the "sample space")
- Purely statistical "sample space" scales
  - The precision with which you know the true effect
    - Power, predictive (conditional) power, P values, posterior probabilities

18

Statistical Tasks

- Understand overall goal
- Refine specific aims (stat hypotheses)
- Materials and methods: Study design
- Collection of data: Advise on QC
- Analysis
  - Describe sample (materials and methods)
  - Analyses to address specific aims
- Interpretation

19

Classification of Statistical Questions

- In order typically used in a new area of science:
  - Clustering of observations
  - Clustering of variables
  - Quantification of distributions
  - Comparing distributions
  - Prediction of individual observations

20

Bottom Line

You better think (think)

about what you're

trying to do…

-Aretha Franklin, "Think"

21

Overall Goal

- "Drug discovery"
  - More generally
    - a therapy / preventive strategy or diagnostic / prognostic procedure
    - for some disease
    - in some population of patients
- A series of experiments to establish
  - Safety of investigations / dose
  - Safety of therapy
  - Measures of efficacy
    - Treatment, population, and outcomes
  - Confirmation of efficacy
  - Confirmation of effectiveness

22

Phases of Investigation

- Series of studies support adoption of new treatment
  - Preclinical
    - Epidemiology including risk factors
    - Basic science:
      - Biochemistry
      - Physiologic mechanisms
      - Physics / engineering
    - Animal experiments: Toxicology / safety
  - Clinical
    - Phase 1: Initial safety / dose finding
    - Phase 2: Preliminary efficacy / further safety
    - Phase 3: Confirmatory efficacy / effectiveness
  - Approval of indication
    - (Phase 4: Post-marketing surveillance, REMS)

23

Phases of Investigation

- A "piecewise continuous" process
- During any individual clinical trial
  - Sequential monitoring, adaptation addresses issues of that trial
- "White space" between trials
  - More detailed analyses
  - Evaluation of multiple endpoints; cost/benefit tradeoffs
  - Exploratory analyses
  - Integration of results from other studies
  - Management decisions
  - Regulatory and ethical review
- Next RCT: May address different question or indication

24

The Enemy

"Let's start at the very beginning, a very good place to start…"

- Maria von Trapp (nee Kutschera)
(as quoted by Rodgers and Hammerstein)

25

First

- Where do we want to be?
  - Describe some innovative experiment?
  - Find a use for some proprietary drug / biologic / device?
    - "Obtain a significant p value"
  - Find a new treatment that improves health of some individuals
    - "Efficacy"
  - Find a new treatment that improves health of the population
    - "Effectiveness"

6

26

Treatment "Indication"

- Disease
  - Therapy: Putative cause vs signs / symptoms
    - May involve method of diagnosis, response to therapies
  - Prevention / Diagnosis: Risk classification
- Population
  - Therapy: Restrict by risk of AEs or actual prior experience
  - Prevention / Diagnosis: Restrict by contraindications
- Treatment or treatment strategy
  - Formulation, administration, dose, frequency, duration, ancillary therapies
- Outcome
  - Clinical vs surrogate; timeframe; method of measurement

27

Scientific Method

- Planned experiment includes protocol specified in advance, including
  - Overall goal
  - Specific aims
  - Materials: Patients, treatments
  - Methods: Administration, monitoring, outcomes
  - Methods: Statistical analysis plan
    - Sampling plan
    - Statistical models for analysis
    - Planned interpretation of spectrum of results

28

Ideal Results

- Goals of "drug discovery" are similar to those of diagnostic testing in clinical medicine
- We want a "drug discovery" process in which there is
  - A low probability of adopting ineffective drugs
    - High specificity (low type 1 error)
  - A high probability of adopting truly effective drugs
    - High sensitivity (low type 2 error; high power)
  - A high probability that adopted drugs are truly effective
    - High positive predictive value
    - Will depend on prevalence of "good ideas" among our ideas

29

14

Frequentist vs Bayesian

- Note both frequentist and Bayesian optimality criteria
  - Sponsor:
    - High probability of adopting beneficial drug (frequentist power)
    - What is probability of continuing investigations? (predictive power)
  - Regulatory:
    - Low probability of adopting ineffective drug (frequentist type 1 error)
    - High probability that adopted drugs work (posterior probability)
- Furthermore, Bayes rule tells us that we can parameterize posterior probability by type 1 error and power
  - Based on prevalence of "good ideas"

$$PPV = \frac{power \times prevalence}{power \times prevalence \quad + \quad type\ I\ err \times (1 - prevalence)}$$

30

Specific Aim

- One of a series of studies used to support adoption of a new standard of treatment
  - Phase 1: Initial safety / dose finding
  - Phase 2: Preliminary efficacy / further safety
  - Phase 3: "Registrational trials"
    - Therapeutics: Establish effectiveness
    - Prevention: Establish efficacy
    - Diagnostics: Establish accuracy
  - Phase 4:
    - Therapeutics: Post-marketing surveillance
    - Prevention: Effectiveness
    - Diagnostics: Impact on outcomes

31

Phase 3 Confirmatory Trials

- The major goal of a "registrational trial" is to confirm a result observed in some early phase study
- Rigorous science: Well defined confirmatory studies
  - Eligibility criteria
  - Comparability of groups through randomization
  - Clearly defined treatment strategy
  - Clearly defined clinical outcomes (methods, timing, etc.)
  - Unbiased ascertainment of outcomes (blinding)
  - Prespecified primary analysis
    - Population analyzed as randomized
    - Summary measure of distribution (mean, proportion, etc.)
    - Adjustment for covariates

32

Why Emphasize Confirmatory Trials?

"When you go looking for something specific, your chances of finding it are very bad, because of all the things in the world, you're only looking for one of them.

"When you go looking for anything at all, your chances of finding it are very good, because of all the things in the world, you're sure to find some of them."

- Darryl Zero in "The Zero Effect"

33

Why Emphasize Confirmatory Trials?

"When you go looking for something specific, your chances of finding [a spurious association by chance] are very bad, because of all the things in the world, you're only looking for one of them.

"When you go looking for anything at all, your chances of finding [a spurious association by chance] are very good, because of all the things in the world, you're sure to find some of them."

34

Multiple Comparisons in Biomedicine

- Observational studies
  - Observe many outcomes
  - Observe many exposures
  - Perform many alternative analyses
    - Summary of outcome distribution, adjustment for covariates
  - Consequently: Many apparent associations
    - May be type 1 errors
    - But even when valid, may be poorly understood due to confounding
- Interventional experiments
  - Exploratory analyses ("Drug discovery")
    - Modification of analysis methods
    - Multiple endpoints
    - Restriction to subgroups

35

Mathematical Basis

- The multiple comparison problem is traced to a well known fact of probability

$$Pr(A \text{ or } B) \geq Pr(A)$$

$$Pr(A \text{ or } B) \geq Pr(B)$$

36

Statistics and Game Theory

- Multiple comparison issues
  - Type I error for each endpoint – subgroup combination
    - In absence of treatment effect, will still decide a benefit exists with probability, say, .025 in each such combination
  - Sequential analysis (more later)
- Multiple endpoints and subgroups increase the chance of deciding an ineffective treatment should be adopted
  - This problem exists with either frequentist or Bayesian criteria for evidence
  - The actual inflation of the type I error depends
    - the number of multiple comparisons, and
    - the correlation between the endpoints

37

Type I Error Inflation: Endpoints, Subgroups

- Experiment-wise error rate from multiple level .05 tests
  - Alternative summary measures are positively correlated
  - Alternative clinical endpoints are usually positively correlated
  - Subgroups defined by the same variable are independent

Number Compared	Worst Case	Correlation 0.00	0.30	0.50	0.75	0.90
1	.050	.050	.050	.050	.050	.050
2	.100	.098	.095	.090	.081	.070
3	.150	.143	.137	.126	.104	.084
5	.250	.226	.208	.184	.138	.101
10	.500	.401	.353	.284	.193	.127
20	1.000	.642	.540	.420	.258	.154
50	1.000	.923	.806	.624	.353	.193

38

Type I Error Inflation: Summary Measures

- Example: Type I error with **normal** data
  - Consider six different summary measures

Any single test:	0.050
Mean, geometric mean	0.057
Mean, Wilcoxon	0.061
Mean, geom mean, Wilcoxon	0.066
Above plus median	0.085
Above plus Pr (Y > 1 sd)	0.127
Above plus Pr (Y > 1.645 sd)	0.169

39

Type I Error Inflation: Summary Measures

- Example: Type I error with **lognormal** data
  - Consider six different summary measures

Any single test:	0.050
Mean, geometric mean	0.074
Mean, Wilcoxon	0.077
Mean, geom mean, Wilcoxon	0.082
Above plus median	0.107
Above plus Pr (Y > 1 sd)	0.152
Above plus Pr (Y > 1.645 sd)	0.192

40

Real-life Examples

- Effects of arrhythmias post MI on survival
  - Observational studies: high risk for death
  - CAST: Specific anti-arrhythmics have higher mortality
- Effects of beta-carotene on lung CA and survival
  - Observational studies: high dietary beta carotene has lower cancer incidence and longer survival
  - CARET: beta carotene supplementation in smokers leads to higher lung CA incidence and lower survival
- Effects of hormone therapy on cardiac events
  - Observational studies: HT has lower cardiac morbidity and mortality
  - WHI: HT in post menopausal women leads to higher cardiac mortality

41

Statistics and Game Theory

- Multiple comparison issues
  - Type 1 error for each endpoint – subgroup combination
    - In absence of treatment effect, will still decide a benefit exists with probability, say, .025 in each such combination
  - Sequential analysis (more later)
- Multiple endpoints and subgroups increase the chance of deciding an ineffective treatment should be adopted
  - This problem exists with either frequentist or Bayesian criteria for evidence
  - The actual inflation of the type 1 error depends
    - the number of multiple comparisons, and
    - the correlation between the endpoints

42

Need for Exploratory Science

- Before we can do a large scale, confirmatory Phase 3 trial, we must have
  - A hypothesized treatment indication to confirm
    - Disease
    - Patient population
    - Treatment strategy
    - Outcome
  - Comfort with the safety / ethics of human experimentation
- In “drug discovery”, in particular, we will not have much experience with the intervention

43

Phase 2 Clinical Trials

- Preliminary evidence of efficacy
  - Goals:
    - Screening for any evidence of treatment efficacy
    - Incidence of major adverse effects
    - Decide if worth studying in larger samples
      - Gain information about best chance to establish efficacy
        - » Choose population, treatment, outcomes
  - Methods
    - Relatively small number of participants
    - Participants closer to true target population
    - Outcome often a surrogate
    - Sometimes no comparison group (especially in cancer)

44

Screening Studies as Diagnostic Tests

- Clinical testing of a new treatment, preventive agent, or diagnostic method is analogous to using laboratory or clinical tests to diagnose a disease
  - Goal is to find a procedure that identifies truly beneficial interventions
- Not surprisingly, the issues that arise when screening for disease apply to clinical trials
  - Predictive value of a positive test is best when prevalence is high
  - Screening trials increase prevalence of beneficial treatments
    - Major predictor of "positive" phase 3 trial is to start with a treatment that is likely to be beneficial

45

Preliminary Studies in Screening

- In cancer less than 5% of treatments studied in clinical trials are adopted
- NCI drug development program 1970 - 1985
  - 350,000 unique chemical structures studied
  - 83 pass preclinical and phase 1 testing
  - 24 pass phase 2 tests for biological activity

46

Preliminary Studies in Screening

- Two general approaches to studying new treatments
- Scenario 1:
  - Study every treatment in a large definitive experiment
    - Only do Phase 3 studies
      - Level of significance 0.025, high power
    - (Ignore, for now, the safety / ethics of this)
- Scenario 2:
  - Perform small screening trials, with confirmatory trials of promising treatments passing early tests
    - Phase 2 studies
      - Level of significance, power (sample size) to be determined
    - Confirmatory Phase 3
      - Level of significance 0.025, high power

47

Scenario 1: Only Phase 3

- Only large trials using 1,000,000 subjects
  - 10% of drugs being investigated truly work
  - Level of significance .025, .025, or 0.05
  - Sample size / power
    - 979 subjects, α=0.025, 97.5% power → 1,021 RCT
    - 500 subjects, α=0.025, 80.0% power → 2,000 RCT
    - 394 subjects, α=0.050, 80.0% power → 2,538 RCT
  - Results
    - N= 979: 99 effective / 23 ineffective (PV+ = .81)
    - N= 500: 160 effective / 45 ineffective (PV+ = .78)
    - N= 394: 202 effective / 114 ineffective (PV+ = .64)

48

Scenario 2a: Screening Phase 2

- Use 700,000 subjects in Phase 2 studies
  - 10% of drugs being investigated truly work at start of phase 2
  - Level of significance .025
  - Sample size / power
    - 100 subjects provide 24% power ➔ 7,000 RCT
  - Results
    - N= 100: 168 effective / 158 ineffective (PV+ = .52)
- Use 300,000 subjects in confirmatory Phase 2 studies
  - 52% of drugs being investigated truly work
  - Level of significance .025
  - Sample size / power
    - 921 subjects provide 96.7% power ➔ 326 RCT
  - Results
    - N= 921: 162 effective / 4 ineffective (PV+ = .98)

49

Scenario 2b: Screening Phase 2

- Use 700,000 subjects in Phase 2 studies
  - 10% of drugs being investigated truly work at start of phase 2
  - Level of significance .10
  - Sample size / power
    - 342 subjects provide 85% power ➔ 2,047 RCT
  - Results
    - N= 342: 173 effective / 184 ineffective (PV+ = .49)
- Use 300,000 subjects in confirmatory Phase 3 studies
  - 49% of drugs being investigated truly work
  - Level of significance .025
  - Sample size / power
    - 839 subjects provide 95% power ➔ 357 RCT
  - Results
    - N= 839: 165 effective / 5 ineffective (PV+ = .97)

50

Summary

		Scenario 1	Scenario 2a	Scenario 2b
Phase 2	Number RCT	2,000 (10% eff)	7,000 (10% eff)	2,047 (10% eff)
	N per RCT	0	100	342
	Type 1 err; Pwr		0.025; 24%	0.100; 85%
	"Positive" RCT		168 eff; 158 not	173 eff; 184 not
Confirmtaory Phase 3	Number RCT	2,000 (10% eff)	326 (52% eff)	357 (49% eff)
	N per RCT	500	921	839
	Type 1 err, Pwr	0.025; 80%	0.025; 97%	0.025; 95%
	# Effctve Adopt	160	162	165
	# Ineff Adopt	45	4	5
	Pred Val Pos	78%	98%	97%
	N per Adopt	500	1,021	1,181

51

Screening Phase 2: Bottom Line

- Pilot studies increase the predictive value of a positive study while using the same number of subjects.
  - Screening parameters can be optimized
    - Proportion of subjects in Phase 2 vs Phase 3
    - Type 1 error at Phase 2
    - Power at Phase 2
- Additional considerations when choosing among screening parameters
  - Will we have same prevalence of "good" ideas when we screen 2,000 drugs vs 7,000 drugs?
  - Holding predictive value of positive constant, which strategy provides more information about safety and secondary endpoints for the treatments eventually adopted?

52

Burden of Larger Phase 2 Studies?

- It appears to be advantageous to use larger Phase 2 studies than is typical currently in cancer research
- BUT: Ethical and efficiency concerns can be addressed through sequential sampling
  - During the conduct of the study, data are analyzed at periodic intervals and reviewed by the DMC
  - Using interim estimates of treatment effect decide whether to continue the trial
  - If continuing, decide on any modifications to
    - scientific / statistical hypotheses and/or
    - sampling scheme

53

Group Sequential Approach

- Perform analyses when sample sizes $N_1 \ldots N_J$
  - Can be randomly determined independent of effect
- At each analysis choose stopping boundaries
  - $a_j < b_j < c_j < d_j$
- Compute test statistic $T_j = T(X_1 \ldots X_{Nj})$
  - Stop if $T_j < a_j$ (extremely low)
  - Stop if $b_j < T_j < c_j$ (approximate equivalence)
  - Stop if $T_j > d_j$ (extremely high)
  - Otherwise continue
- Boundaries chosen to protect experimentwise
  - Type 1 error
  - Power

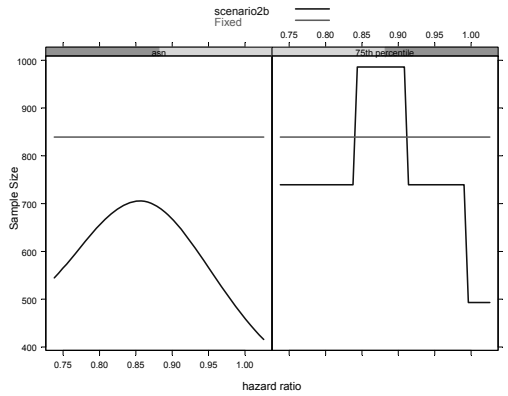
54

Potential Benefits of Stopping Rules

- Sequential sampling
  - Aggressive stopping at small N for futility: Pocock boundaries
    - Greatest efficiency (or nearly so)
  - Conservative stopping at small N for efficacy: O'Brien-Fleming
    - Burden of proof, other endpoints
- Type 1 error, power maintained exactly at each phase
  - Worst case maximum sample size increases
- Average sample size requirements assuming 10% truly effective drugs at start of Phase 2
  - Only large studies : 58.5% of fixed sample
  - Pilot scenario 2a : 56.0%
  - Pilot scenario 2b : 61.0%

55

Scenario 2b Average Phase 3 Sample Size



56

Furthermore

- Additional advantages of screening trials
  - Gathering more detailed preliminary safety data before embarking on expensive, large scale Phase 3 trials
  - Gathering preliminary efficacy data that allows fine tuning of
    - Fine tune eligibility criteria
      - Include only susceptible patient populations
      - Exclude patients at high risk for AEs
    - Optimal treatment strategies
      - Fine tune formulation, dose, administration, frequency, duration
      - Develop dose modification strategies
      - Prophylactic treatments, rescue treatments for AEs
    - Optimal clinical endpoints
- Major disadvantage
  - "White space" (time delay) between phase 2 and phase 3

57

Phase 2 Clinical Trials: Methods

- As typically implemented, the screening role of Phase 2 RCT is somewhat attenuated
  - Disease definition may be restrictive due to desires for
    - Efficacy: to demonstrate proof of concept
    - Safety / ethics: Caution with the unknown treatment in extremely serious disease (e.g., cancer)
  - Participants may not be true target population
    - Heavier trial burden in early trials
    - Unknown impact of concomitant disease
  - Outcome often a surrogate
    - Reduce costs / duration of RCT
    - Plausibility of effect on clinical outcome

58

Inflation of the Type 1 Error

- Recall that in order to avoid inflation of type 1 error, we require confirmatory studies using prespecified
  - Patient population
  - Treatment
  - Primary clinical outcome
  - Statistical analysis
- Hence, we must be concerned about data dredging ("data mining") of the phase 2 data, because it may lead to differences between phase 2 and phase 3 due to
  - Revising outcomes to reflect the most promising results
  - Revising eligibility criteria based on subgroup analyses
  - Changing from surrogate efficacy to effectiveness endpoints
    - "Treating the symptom not the disease"

59

Generalizability

- The previous results are dependent on
  - A mixture of 10% effective drugs and 90% ineffective drugs, *where "effectiveness" is defined based on the clinical outcome used in the phase 3 trial*
  - Phase 2 and phase 3 type 1 errors being controlled at the specified level *based on the phase 3 outcome*
  - Phase 2 and phase 3 power being controlled at the specified level *based on the phase 3 outcome*
- If a surrogate outcome is used at phase 2, the "positive results" may not generalize to the true clinical outcome used at phase 3

60

Strategies

- Settings
  - Homogeneous effect: works equally well in all subgroups
  - Inhomogenous effect: only works in half of population
  - True outcome vs misleading surrogate
- Strategies
  - Prespecified analysis with control of type 1 error, power
    - Optimal choice would depend on prior belief of prevalence
  - Adaptive modification of endpoints or subgroups ("enrichment")
    - Data dredging, no control of type 1 error, power
    - Adaptive designs controlling type 1 error, power
      - Fully adaptive using conditional error, weighted statistics, etc.
      - Prespecified using sufficient statistic

61

Comparisons

	RCT	Eff	Not	n
• Nonadaptive				
– Homogeneous effect	2,047	165	5	1,181
– Homogeneous, 10% misleading	1,812	147	8	1,181
– Homogeneous, 20% misleading	1,627	132	12	1,181
– Inhomogeneous effect	2,123	99	5	1,181
• Adaptive subgroups: inflate error				
– Homogeneous effect	1,485	134	11	1,181
– Inhomogeneous effect	1,490	109	11	1,181
• Adaptive subgroups: control error				
– Homogeneous effect	1,707	139	4	1,277
– Inhomogeneous effect	1,720	105	4	1,277

62

Bottom Line

- Screening Phase 2 trials provide great protection
  - Ensure that overwhelming majority of adopted therapies are truly effective
- Control of type 1 and 2 errors is important at phase 2
  - But note that type 1 error of 0.025 not necessarily indicated
- Adaptive designs can help provide that control
  - But need to re-power the study to get greatest benefit
  - The added benefit over nonadaptive designs is not huge, but there are advantages
    - Higher power and predictive value of the positive
    - More beneficial drugs identified
    - More patient exposure for adopted drugs
- Adaptation cannot protect against false surrogates

63