Protocol Development: Study Design

November 6, 2006
Elizabeth Garrett-Mayer (esg@jhu.edu)
Jane Perlmutter (janep@gemini-grp.com)
Objectives

1. Obtain a general appreciation of how trial design can impact patient accrual and trial costs, as well as scientific validity.

2. Clarify issues and potential solutions to incorporating biomarkers into clinical research and practice.

3. Become familiar with a variety of innovative trial designs.

4. Practice articulating provocative questions about proposed trial designs.

5. Be aware of statistical and trial designs issues currently being considered by the FDA.
Topics

- Overview of Protocol Development
- Impacts of Trial Designs
- Promise and Challenge of Biomarkers
- Correlative Science
- Multi-Stage/Phase Designs
- Bayesian vs. Traditional (Frequentist) Statistics
- Adaptive and Other Innovative Designs
- FDA’s Statistical and Trial Design Initiatives
Group Assessment

How much experience have you had with trial protocols?

- Hardly any knowledge—e.g., I tune out of discussions and never read protocols
- Limited—e.g., I’ve has a bit of training or experience, but am not at all comfortable dealing with these issues
- Fair—e.g., Although I know I miss some of the subtleties, I feel fairly comfortable reviewing protocols
- Good—e.g., I am quite interested and/or comfortable with reviewing protocols
Protocol Development Process

CC = Cooperative Committee
CG = Cooperative Group
CTEP = NCI’s Cancer Therapy Evaluation Program
IRB = Institutional Review Board
Protocol Components

- Introduction / Background
- Study Hypothesis
- **Study Design**
- Rationale / Study Objectives
- Eligibility
- **Preliminary Statistical Design**
- Competing Protocols
- Feasibility
- Schema / Flowchart
- Informed Consent Process
- Communication Plan
- Dealing with Results
Topics

- Overview of Protocol Development
- Impacts of Trial Designs
- Promise and Challenge of Biomarkers
- Correlative Science
- Multi-Stage/Phase Designs
- Bayesian vs. Traditional (Frequentist) Statistics
- Adaptive and Other Innovative Designs
- FDA’s Statistical and Trial Design Initiatives
How would you assess your current level of knowledge of statistics and clinical trial design?

- Hardly any knowledge—e.g., I tune out of discussions and skip reading sections that address these issues
- Limited—e.g., I’ve has a bit of training, but am not at all comfortable dealing with these issues
- Fair—e.g., Although I know I miss some of the subtleties, I feel fairly comfortable addressing these issues
- Good—e.g., I am quite interested and/or comfortable with this issues
Innovative Trial Design: Why?

- Changes in science require it
- Innovations in trial design can:
  - Make trial them more attractive to patients
  - Reduce number of patients and funding required
  - Speed conclusions
  - Improve inferences
- Innovations in trial design have high leverage, because they can be applied to any disease or treatment
What Changes in Science are Challenging Trial Design?
What Changes in Science are Challenging Trial Design?

- Realization that cancer is many different diseases
- Increasing interest in correlative science
- Time to events is often of more interest than response rate
- As treatments improve, “events” become less frequent
- Increasing interest in simultaneously looking at multiple independent variables (e.g. patient characteristics, treatment combinations and delivery schedules) and dependent variables (e.g., tumor progression, side effects, QOL)
- Paradigm for dose finding should be different for non-cytotoxic agents
Topics

- Overview of Protocol Development
- Impacts of trial designs
- **Promise and challenge of biomarkers**
- Correlative Science
- Multi-Stage/Phase Designs
- Bayesian vs. Traditional (Frequentist) statistics
- Adaptive and Other Innovative Designs
- FDA’s Statistical and Trial Design Initiatives
What Are All of These?

- Surrogate Endpoint
- Primary Endpoint
- Surrogate Biomarker
- Biomarker
- Secondary Endpoint
- Gene Array
Definitions

- **Primary Endpoint:**
- **Secondary Endpoint:**
- **Surrogate Endpoint:**
- **Biomarker:**
- **Gene Array:**
- **Surrogate Marker:**
How Are Biomarkers be Used?
How Are Biomarkers be Used?

- Risk assessment
- Screening
- Differential diagnosis
- Prognosis
- Prediction
- Monitor disease

Can Biomarkers be Considered “Surrogate Markers?”

- Replace a distal endpoint (e.g. survival) by proxy endpoint (e.g. metabolic tissue activity).

- Benefits of using surrogate markers
  - Reduction in sample size
  - Reduction in trial duration
  - Reduction in cost
  - Reduction in time to evaluate new therapies
  - Potential benefits to the patient: earlier intervention when initial therapy is not effective

- Their use is NOT AS EASY AS IT SOUNDS…

- Use of a marker as surrogate for outcome requires that you first identify one.
What is a Surrogate Biomarker?

- **Defining Characteristic:**
  - a marker must predict clinical outcome, in addition to predicting the effect of treatment on clinical outcome

- **Operational Definition**
  - establish an association between marker & clinical outcome
  - establish an association between marker, treatment & clinical outcome, in which marker mediates relationship between clinical outcome and treatment
1) establish an association between biomarker & clinical outcome.

2) establish an association between biomarker, treatment & clinical outcome, in which biomarker completely mediates relationship between clinical outcome and treatment.
NOT Surrogate Markers

- Treatment
- Clinical outcome
- Biomarker
- Treatment
- Clinical outcome
- Biomarker
Marker Studies

How are marker studies different than other studies?

- Primary outcomes are ‘efficacy-related’, but not clinical.
- Outcomes are ‘surrogate’ outcomes or ‘correlative’ outcomes.
- Measuring these outcomes is often more invasive and more costly than standard safety or efficacy trials.
- Measurement of these outcomes can be complicated.
Early PET: Surrogate Marker?

<table>
<thead>
<tr>
<th></th>
<th>Early</th>
<th>Late</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+</td>
<td>-</td>
<td>total</td>
</tr>
<tr>
<td>Early</td>
<td>+</td>
<td></td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>13</td>
<td>47</td>
</tr>
<tr>
<td>total</td>
<td>33</td>
<td>47</td>
<td>80</td>
</tr>
</tbody>
</table>

Sensitivity = 20/20 = 100%
Specificity = 47/60 = 78%

PPV = 20/33 = 61%
NPV = 47/47 = 100%

Early PET: Surrogate Marker?

<table>
<thead>
<tr>
<th>CR+PR</th>
<th>+</th>
<th>-</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>21</td>
<td>44</td>
<td>65</td>
</tr>
<tr>
<td>-</td>
<td>15</td>
<td>8</td>
<td>23</td>
</tr>
<tr>
<td>total</td>
<td>36</td>
<td>52</td>
<td>88</td>
</tr>
</tbody>
</table>

Sensitivity = 15/23 = 65%
Specificity = 44/65 = 67%

PPV = P(CR+PR negative | early PET positive) = 15/36 = 42%
NPV = P(CR+PR positive | early PET negative) = 44/52 = 85%
Early PET: Surrogate Marker?

Event = death, progression, or relapse.

<table>
<thead>
<tr>
<th>Event</th>
<th>+</th>
<th>-</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event +</td>
<td>16</td>
<td>6</td>
<td>22</td>
</tr>
<tr>
<td>Event -</td>
<td>20</td>
<td>48</td>
<td>68</td>
</tr>
<tr>
<td>Total</td>
<td>36</td>
<td>54</td>
<td>90</td>
</tr>
</tbody>
</table>

Sensitivity = 16/22 = 73%
Specificity = 48/68 = 71%

PPV = P(event | early PET positive) = 16/36 = 44%
NPV = P(no event | early PET negative) = 48/54 = 89%
Dramatic, but True Surrogate?

**A** Event-Free Survival According to Response at 2 Cycles on the Basis of PET (n=90)

- PET (-), n=54
- PET (+), n=36

Probability of Event-Free Survival

Years After Randomization

\[ P < .001 \]

**B** Overall Survival According to Response at 2 Cycles on the Basis of PET (n=90)

- PET (-), n=54
- PET (+), n=36

Probability of Overall Survival

Years After Randomization

\[ P = .006 \]
Surrogate Markers

- So, what have you learned?

- If early PET provides a **NEGATIVE** result, how would the result affect the course of action?

- If early PET provides a **POSITIVE** result, how would the result affect the course of action?

- Put another way (more statistically): if there are two patients who are the same in every way (IPI, performance status, age, etc.), but one had a positive PET at 4 weeks and the other had a negative PET at 4 weeks, would these results be convincing enough for you to act differently?
What Do You Need to Know About Biomarkers?
What Do You Need to Know About Biomarkers?

- What evidence is required to validate that a biomarker assay does as it claims?
- What evidence is required to validate the use of a biomarker in clinical practice?
- What are some special issues associated simultaneously developing drugs and biomarkers?
- What are some special issues associated with multi-gene tests?
- What are some special issues associated with using banked tissue in prospectively designed trials of biomarkers?
- What are some special issues associated with doing good correlative science in the context of clinical trials?
- What are some of the ethical issues raised by biomarkers?
Topics

- Overview of Protocol Development
- Impacts of Trial Designs
- Promise and Challenge of Biomarkers

Correlative Science
- Multi-Stage/Phase Designs
- Bayesian vs. Traditional (Frequentist) Statistics
- Adaptive and Other Innovative Designs
- FDA’s Statistical and Trial Design Initiatives
Group Assessment

In the context of the cooperative groups, correlative studies:

- Are generally supplemental to the main issues of a clinical trial
- Are “hypothesis generating” (sometimes referred to as fishing expeditions)
- Are typically underpowered
- All of the above
Correlative Studies

- It is still valid to look at markers that you would expect to be correlated with the clinical outcome.
- But, we do not want to be overconfident by saying that they are true ‘surrogates.’
- Correlative studies might include:
  - Pharmacokinetics
  - Pharmacodynamics
  - Imaging
  - Other biologic markers that can be measured in serum, biopsy samples, etc.
Correlative Studies

- Design can be tricky
- Since these questions are not of PRIMARY interest, typical designs are not perfectly suited to looking at correlates
- We end up doing the best we can
Choosing the Timing

- **Standard Phase I**
  - Patients are under surveillance for short term toxicities.
  - Patients contact study team when a toxicity occurs after discharge.
  - Not an issue of ‘when to look.’

- **Add in correlative outcomes:**
  - Usually a pre-post setting: how does the measure compare after treatment to before treatment.
  - Baseline (before treatment) measure is needed.
  - Post-treatment measures:
    - When is treatment at its most potent?
    - How often should response be measured?
    - Expensive? Invasive?
    - Is it sufficient to look at clinic visit times?
Phase II/III Designs

- Phase II designs: usually efficacy
- Standard efficacy outcomes
  - Complete response
  - Partial response
  - Overall survival
- For cytostatic agents
  - Progression
  - Progression-free survival
  - Laboratory outcomes…
Defining Outcomes in Laboratory Studies

- They are often messy (e.g. cell counts, gene expression)
- More common binary outcomes have nice properties:
  - “Looking for 40% response vs. 20% response”
- Laboratory outcomes are not so nice:
  - Often skewed.
  - Often have ‘undetectable’ range.
  - Often do not know what to expect.
  - This makes it hard to plan (i.e. sample size, power).
- Novel assays: Not obvious what expected changes would be without treatment.
  - How much fluctuation would we expect to see?
Expected Fluctuations

But, with multiple patients, these changes would tend to average to zero.

However, if half of patients have increase and half have decrease, we might conclude that treatment is 50% effective!
Follow-up times are compared to baseline
- That puts a lot of stock in baseline measure
- Why not consider a ‘burn-in’ period?

If baseline is inaccurately measured, all comparisons will be incorrect.
Measurement Issues

- Regardless of natural fluctuations...
- How accurately are we measuring our outcome?
- How accurate can we measure blood flow?
- Are we on target?
Measurement Issues

- Why would it be measured inaccurately?
  - Often LONG protocol to get ‘final’ measure.
  - Lots of room for errors!

- Although the classic efficacy outcome of ‘response’ is relatively soft, it has been objectively defined.

- Laboratory endpoints require assays and other measurements, and also assumptions.

- Often, assay is being developed along with the trial.
Assay Properties

- **IS THE ASSAY SENSITIVE?**
- **Sensitivity**: does the assay detect abnormalities in cases where abnormalities exist?
- **Specificity**: does the assay find normal levels in normal cases?
- “It has been shown that at high flow rates, measuring blood flow by PET underestimates blood flow.”
  - bias in results.
More Measurement Issues

- Reliability of assay
  - How reproducible are the results?
    - Two samples taken from the same patient on the same day from different lesions?
    - Two samples taken from the same patient on the same day from the same lesion?
    - One sample analyzed twice using the same method?
  - Subjectivity
  - Inter-rater agreement
  - Intra-rater agreement
  - In what ways can ‘error’ come into the procedure?

Great study + bad assay = bad study
Measurement Issues

- How can we know the reliability of the measures?
- Preliminary studies (pre-clinical).
- Build it into the study design!
  - Reliability substudy:
    - Inter-rater agreement
    - Intra-rater agreement
  - Incorporate burn-in period
  - Take multiple measures
  - Run the assay (test) more than once
Trials looking at surrogate and correlative outcomes often need creative/novel study designs.

Understanding the properties (i.e. reliability, sensitivity) of your assay/measurement technique are crucial.

Think carefully about:

- Are your markers truly surrogates?
- Timing
- The potential for missing data, especially in biopsy studies (Jackie Walling)…..
Practical Issue: Biopsies in Clinical Trials

- Including biopsies is great for research
  - Allows investigation of correlative endpoints
  - Helps understand mechanism of action

- But, practically:
  - Often hurts accrual
  - Mandatory versus optional?
  - Potentially large proportion are ‘unevaluable’
  - Might only be useful if both of paired samples are evaluable
  - Is it worth the effort (and is it ethical) if you only end up with useable information on a subset of the patients?
Assessability of ex vivo samples used in chemosensitivity assays
(from Shrag et al 2004)

<table>
<thead>
<tr>
<th>Ref</th>
<th>Assay</th>
<th>N</th>
<th>% pts with assessable assays</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>HTCA</td>
<td>470</td>
<td>64</td>
</tr>
<tr>
<td>2</td>
<td>HTCA</td>
<td>168</td>
<td>45</td>
</tr>
<tr>
<td>3</td>
<td>CCS</td>
<td>68</td>
<td>71</td>
</tr>
<tr>
<td>4</td>
<td>DiSC</td>
<td>79</td>
<td>33</td>
</tr>
<tr>
<td>5</td>
<td>DiSC</td>
<td>45</td>
<td>78</td>
</tr>
<tr>
<td>6</td>
<td>Mod DiSC</td>
<td>SCLC:115 NSCLC:165</td>
<td>29</td>
</tr>
<tr>
<td>7</td>
<td>MTT</td>
<td>83</td>
<td>88</td>
</tr>
<tr>
<td>8</td>
<td>DiSC</td>
<td>18</td>
<td>56</td>
</tr>
<tr>
<td>9</td>
<td>ATP</td>
<td>31</td>
<td>93</td>
</tr>
<tr>
<td>10</td>
<td>EDR</td>
<td>50</td>
<td>100</td>
</tr>
</tbody>
</table>
Opinion Leader Survey: Assessability in Paired Samples

- 17 OLs at Major Cancer Centers in the US and UK
- 2 phase I studies, 1 phase I/II, 12 phase II studies, largely ongoing
- 11/15 studies mandatory bx

<table>
<thead>
<tr>
<th>Number of patients recruited (range)</th>
<th>9-58</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Range of paired mandatory biopsies</td>
<td>43 -100</td>
</tr>
<tr>
<td>% of Patients with paired evaluable biopsies in mandatory studies</td>
<td>27 - 61</td>
</tr>
<tr>
<td>% of Patients with paired evaluable biopsies in optional studies</td>
<td>10 - 58</td>
</tr>
</tbody>
</table>
OL Survey: Take Home Messages

- Paired biopsy studies are hard!
  - No center has this cracked

- Multiple issues:
  - Clinical Trial Design
  - Institutional
  - Technical
  - QC
    - Unknown false positives/negatives
  - Assays
    - Few / no correlations with xenograft / pk data/clinical data
    - Standardization/ Quantitation
  - Ethics
Topics

- Overview of Protocol Development
- Impacts of Trial Designs
- Promise and Challenge of Biomarkers
- Correlative Science
- Multi-Stage/Phase Designs
- Bayesian vs. Traditional (Frequentist) Statistics
- Adaptive and Other Innovative Designs
- FDA’s Statistical and Trial Design Initiatives
What Questions Should Advocates Ask About Trial Designs?
What Questions Should Advocates Ask About Trial Designs?

- Is there a way to design this trial so that fewer patients are needed?
- Is there a way to design this trial so that fewer patients exposed to an ineffective treatment?
- Is there a way to design this trial so that we can more rapidly determine the results?
- Is the trial likely to accrue a diverse group of patients? Are there design changes that are likely to increase overall and/or minority accrual?
- Can the design be modified to not only determine whether the treatment has benefit, but also who is likely to benefit?
- Is the design powered to detect clinically significant findings?
The “Gold Standard”

**Two Arm Randomized Clinical Trial**

<table>
<thead>
<tr>
<th>Experimental Treatment</th>
<th>Standardized Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Equal numbers of patients randomly assigned to two groups
- Total number of patients planned to achieve a given level of statistical power to detect a difference of a pre-specified level
- Single endpoint—simple comparison
- May require many trials to answer complicated questions
Two-Stage Designs
(Especially Useful for Phase 2)

- If by the 15\textsuperscript{th} patient you’ve seen no responses, is it worth proceeding?
- Maybe you should have considered a design with an early stopping rule
- Two-stage designs:

\begin{itemize}
  \item Stage 1: enroll $N_1$ patients
  \item $X_1$ or more respond
  \item Fewer than $X_1$ respond
\end{itemize}

\begin{itemize}
  \item Stage 2: Enroll an additional $N_2$ patients
  \item Stop trial
\end{itemize}
Revised Design

- **Stage 1**: enroll 19 patients
  - If 4 or more respond, proceed to stage 2
  - If 3 or fewer respond, stop

- **Stage 2**: enroll 20 more patients (total N=39)
  - If 12 or more of total respond, conclude effective
  - If 11 or fewer of total respond, conclude ineffective

- Design properties?

  \[
  \alpha = 0.10 \\
  \beta = 0.10 \text{ (power } = 0.90) \\
  H_0 : p = 0.20 \text{ (null response rate)} \\
  H_1 : p = 0.40 \text{ (target response rate)}
  \]
Simon Two-Stage Designs

- MANY designs fit the criteria
- “Optimal”
  - Minimum *expected* sample size under $H_0$
  - Minimum *maximum* sample size
- Preserves alpha and power, and permits early look
Early Stopping

- FUTILITY stopping
- Simon design ONLY allows stopping early if there is strong evidence that the treatment is not efficacious
- Can also have early stopping for efficacy
  - Generally not popular
  - Important to accumulate evidence to support claim of efficacy
  - But, not stopping prolongs time to launch phase III
Multiple Outcomes

- Phase II = “safety + efficacy” trial
- Then why are we only talking about efficacy?
- Bryant and Day (1995): extend Simon two-stage to incorporate both outcomes
  - Early stopping if either efficacy OR futility is not sufficiently high
  - Requires both safety and efficacy to meet a threshold before continuing to stage 2
  - Same idea as Bryant and Day but more elegant mathematically
Another “novel” issue

- Time to event outcomes in Phase II
  - Response rate no longer the ‘outcome of choice’ in Phase II studies
    - targeted agents may not shrink cancer
    - we’re learning: tumor shrinkage ≠ increased survival
  - Time to event outcomes more common
    - time to progression
    - time to relapse
  - More than ever, need early stopping
  - Simon’s two-stage does not apply
  - Bayesian and Likelihood methods are becoming more appealing
Phase II/III studies

- Several versions {Schaid (1988), Storer (1990), Ellenberg and Eisenberger (1985), Scher and Heller (2002)}

- General idea
  - Begin with randomized phase II study
  - Randomize to control arm & experimental arm(s)
  - If some threshold of efficacy is met, continue to phase III sample size for direct comparison

- Benefits:
  - Allow use of phase II data in phase III inference
  - Minimize delay in starting up phase III study
  - Uses concurrent control
Phase II/III studies

Cons:
- The sample size for the phase II part is approximately twice as large as would be needed for standard phase II.
- Need phase III infrastructure developed even if it stops early.
- Would be useful if MOST phase II studies showed efficacy (not the case!)
- Really, these could be considered phase III designs with very aggressive early stopping rules.
Topics

- Overview of Protocol Development
- Impacts of Trial Designs
- Promise and Challenge of Biomarkers
- Correlative Science
- Multi-Stage/Phase Designs
- Bayesian vs. Traditional (Frequentist) Statistics
- Adaptive and Other Innovative Designs
- FDA’s Statistical and Trial Design Initiatives
Inferential Statistics 101 Questions

- **Bayesian Statistics:**
  a. Is a new approach to analyzing data
  b. Incorporate prior knowledge into statistical inference
  c. Both a and b
  d. Neither a nor b

- **If there are no statistically significant results:**
  a. The treatments are the same
  b. The treatments are different
  c. The treatments may be the same or different
  d. Other
Inferential Statistics 101 Questions

■ The null hypothesis is:
  a. What you hope is true
  b. What you think is true
  c. What you assume is true
  d. Other

■ In hypothesis testing you:
  a. Determine how likely the null hypothesis, given your data
  b. Determine how likely your data, given the null hypothesis
  c. Both a and b
  d. Neither a nor b
Introduction to Bayesian Approach

- An alternative approach to statistical inference
- Based on continuous learning; allows incorporation of prior knowledge
  - Start with prior information (“prior distribution”)
  - Observe data (“likelihood function”)
  - Combine prior and data to get “posterior” distribution
  - Make inferences based on posterior
- Controversial, but increasingly being accepted
- Sets the stage for , although not strictly required for adaptive trials
Traditional Approach to Statistical Inference

Reject or Fail to Reject $H_0$

$H_0$

$Pr(\text{Data} | H_0)$ (or more extreme)

“In Hypothetical” Component

“Data” Component i.e. current experiment
Bayesian Approach to Statistical Inference

\[
\frac{Pr(H_0 \mid Data)}{Pr(H_1 \mid Data)} = \frac{Pr(H_0)}{Pr(H_1)} \times \frac{Pr(Data \mid H_0)}{Pr(Data \mid H_1)}
\]

**Post-test Odds** = **Pre-test Odds** \(\times\) **Likelihood Ratio**

*a.k.a. Bayes factor*

---

**Inference (probability)**

*e.g. prior results, theoretical basis*

---

**“Subjective” Component**

---

**“Data” Component**

*i.e. current experiment*
Bayesian Inference

- No p-values and confidence intervals
- Designs are not quantified by $\alpha$ and $\beta$
- From the posterior distribution:
  - Posterior probabilities
  - Prediction intervals
  - Credible intervals
- Bayesian designs
  - Can look at data as often as you like (!)
  - Use information as it accumulates
  - Make “what if?” calculations
  - Helps decide to stop now or not
Bayesian Designs

- Requires ‘prior’
  - Reflects uncertainty about the response rate
  - Can be ‘vague’, ‘uninformative’
  - Can be controversial: inference may change
Posterior Probabilities

Cumulative Number of Patients

Probability

- Probability $p < 0.20$
- Probability $p > 0.40$
Other priors

- What if we had used a different prior?
- Assume informative “orange” prior
## Traditional vs Bayesian Methods

<table>
<thead>
<tr>
<th>Issue</th>
<th>Traditional Methods</th>
<th>Bayesian Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prior information other than in the study being analyzed</strong></td>
<td>• Informally used in design</td>
<td>• Used formally by specifying a prior probability distribution</td>
</tr>
<tr>
<td><strong>Interpretation of the parameter of interest</strong></td>
<td>• A fixed state of nature</td>
<td>• An unknown quantity which can have a probability distribution</td>
</tr>
<tr>
<td><strong>Basic question</strong></td>
<td>• “How likely is the data, given a particular value of the parameter?”</td>
<td>• Plot of posterior distributions of the parameter</td>
</tr>
<tr>
<td></td>
<td>• Hypothesis testing</td>
<td>• Calculation of specific posterior probabilities of interest</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Use of the posterior distribution in formal decision analysis</td>
</tr>
<tr>
<td><strong>Presentation of results</strong></td>
<td>• P values and estimates adjusted for the number of analyses</td>
<td>• Inference not affected by the number or timing of interim analyses</td>
</tr>
<tr>
<td><strong>Interim analyses</strong></td>
<td>• Conditional power analyses</td>
<td>• Predictive probability of getting a firm conclusion</td>
</tr>
<tr>
<td><strong>Dealing with subsets in trials</strong></td>
<td>• Adjusted P values (e.g., Bonferroni)</td>
<td>• Subset effects shrunk towards zero by “skeptical” prior</td>
</tr>
</tbody>
</table>

## Strengths & Weaknesses

<table>
<thead>
<tr>
<th>Weaknesses of Traditional Approach</th>
<th>Strengths of Bayesian Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tends to answer the wrong question</td>
<td>Answers the right questions and agree with natural common sense</td>
</tr>
<tr>
<td>Likelihood principle often violated</td>
<td>Likelihood principle is consistently followed</td>
</tr>
<tr>
<td>Output is very useful for decision making</td>
<td>Output is ideal for decision making</td>
</tr>
<tr>
<td>Too easy to apply mindlessly</td>
<td>Forces careful thought of model and prior probabilities</td>
</tr>
<tr>
<td>Subjective component is hidden</td>
<td>Subjective component is more thorough and transparent</td>
</tr>
<tr>
<td>Analysis is often simplistic and can be misleading</td>
<td>Allows formal incorporation of relevant information other than data immediately at hand</td>
</tr>
<tr>
<td>Overuse of hypothesis testing framework</td>
<td>Lend themselves to messy, multiple data sets (e.g., meta-analysis)</td>
</tr>
</tbody>
</table>

## Challenges Raised by Bayesian Approaches

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Computationally intractable</td>
<td>• Monte Carlo simulation</td>
</tr>
<tr>
<td>• Subjectivity associated with prior Probabilities</td>
<td>• New software</td>
</tr>
<tr>
<td>• Most people were taught and are comfortable with traditional approaches</td>
<td>• Multiple scenarios &amp; Sensitivity Analysis</td>
</tr>
<tr>
<td>• Ambiguity about FDA &amp; journal editors acceptance of non-traditional methods</td>
<td>• Uniform Priors</td>
</tr>
<tr>
<td></td>
<td>• Education</td>
</tr>
<tr>
<td></td>
<td>• This is changing</td>
</tr>
</tbody>
</table>
Topics

- Overview of Protocol Development
- Impacts of Trial Designs
- Promise and Challenge of Biomarkers
- Correlative Science
- Multi-Stage/Phase Designs
- Bayesian vs. Traditional (Frequentist) statistics
- Adaptive and Other Innovative Designs
- FDA’s Statistical and Trial Design Initiatives
Example Designs

- Designs that can improve efficiency
  - Adaptive Patient Assignment
  - “Continuous” Adaptive Design
  - Seamless Phasing
  - Patient Enrichment Strategies

- Designs that may be especially attractive to patients
  - Patient Chooses Design
  - Randomized Discontinuation Design
Group Assessment

**In Adaptive Trials you can:**

- Modify the design after you see the results
- Modify the design as the data accrue, based on rules determined before the design started
- Both a and b
- Neither a nor b
Adaptive Designs

- **Multi-stage Designs:** later stages based on interim results

- **Example Adaptation Rules:**
  - *Allocation Rule:* How are patients allocated to treatment arms? (Note: patients are always randomly assigned, but the relative frequency may be changed, including adding or dropping arms)
  - *Sampling Rule:* How many subjects should be sampled at the next stage? (Note: this may change due to surprises about accrual rate, sample variance, etc.)
  - *Stopping Rule:* When should the study be terminated due to observed efficacy, harm, futility, or safety
Patient Allocation Adaptive Design

- **Difficulties with Traditional Approaches**
  - Trials require many patients, take too long, and are too costly
  - Half of patients in the trial do not receive optimal treatment

- **Potential Solution**

  - Randomly & Equally Assign Patient
  - Observe & Predict Responses
  - Randomly & Unequally Assign Patients
  - True Treatment Effect?
    - Yes
    - No

- If apparent treatment effect is true, groups will diverge & trial can be rapidly completed
- If apparent treatment effect is random, groups will converge
Adapt the randomization to learn while effectively treating patients on trial:

(1) Begin by randomizing with equal chance per arm
(2) Then, adjust probability of assignment to reflect the knowledge of the best treatment
Adaptive Designs

- Begin assuming equally effective (1/3, 1/3, 1/3)
- May wait until a minimum number have been treated per arm
- Based on currently available (accumulated) data, randomize next patient (i.e., “weighted” randomization)
- Stopping rules: drop an arm when there is “strong” evidence that
  - It has low efficacy OR
  - It has lower efficacy than competing treatments
Adaptive Designs

**Summary of trial results:**
- TI dropped after 24th patient
- Trial stopped after 34 patients (TA dropped)

<table>
<thead>
<tr>
<th></th>
<th>Complete responses by 50 days</th>
</tr>
</thead>
<tbody>
<tr>
<td>IA</td>
<td>10/18 = 56%</td>
</tr>
<tr>
<td>TA</td>
<td>3/11 = 27%</td>
</tr>
<tr>
<td>TI</td>
<td>0/5 = 0%</td>
</tr>
</tbody>
</table>
Patient enrichment refers to:

- Restricting participation in trials to patients most likely to benefit from the investigational treatment
- Pre-treating patients with drugs that are intended to build their white cell count
- Providing monetary compensation to patients participating in clinical trials
- None of the above
<table>
<thead>
<tr>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enrichment Designs</td>
<td></td>
</tr>
</tbody>
</table>

**Strengths & Weaknesses of Patient**
Example Patient Enrichment Design
What Design Issues Make Trials Less Unattractive to Patients?
What Design Issues Make Trials Less Unattractive to Patients?

- Randomization
- Not necessarily getting the investigational treatment
- Study procedures (e.g., tests) that are perceived to be more for the benefit of science than their own benefit
- More frequent doctor/clinic visits
What Design Issues Lead to Increased Costs?
What Design Issues Lead to Increased Costs?

- Large sample sizes
- Heterogeneous patient groups, only some of which respond to treatment
- Lack of patient compliance or completion
- Slow patient accrual
Patient Preference Design
(“Out of the Box” Design)

- **Difficulties with Traditional Approaches**
  - Patient accrual is slow
  - <50% of eligible patients who are offered trials actually enroll
  - Some patients are uncomfortable with random assignment

- **Potential Solution**
  - If the direction of any treatment effect is independent of whether or not patients are randomized, fewer randomized patients are required to achieve same power
  - Any “patient-selection” findings may themselves prove interesting

<table>
<thead>
<tr>
<th>Patient Preference Design</th>
<th>Experimental Treatment</th>
<th>Standard Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomized Treatment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient Selected Treatment</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Randomized Discontinuation Design

**Difficulties with Traditional Approaches**
- Trials take too long and are too costly
- Only a small subset of patients typically respond to new drugs

**Potential Solution**

**Randomized Discontinuation Design**

- Initially all patients receive the experimental treatment
- Superiority is based on patients who are initially stabilized with the experimental treatment

Topics

- Overview of Protocol Development
- Impacts of Trial Designs
- Promise and Challenge of Biomarkers
- Correlative Science
- Multi-Stage/Phase Designs
- Bayesian vs. Traditional (Frequentist) statistics
- Adaptive and Other Innovative Designs
- FDA’s Statistical and Trial Design Initiatives
Recent FDA Initiatives

- **Critical Path Initiative**—including:
  - Better Evaluation Tools—Developing New Biomarkers and Disease Models
  - Streamlining Clinical Trials

- **2006 Guidance Documents**
  - Bayesian Statistics in Medical Device Clinical Trials
  - In Vitro Diagnostic Multivariate Index Assays

- **Drug Development Science**: Obstacles and Opportunities for Collaboration Among Academia, Industry and Government
Ongoing FDA Initiatives

- Reviewing existing and current uses of adaptive trials
- Five promised guidance documents:
  - Looking at multiple endpoints in the same trial
  - Using enrichment designs
  - Non-inferiority trial designs
  - Adaptive designs
  - Dealing with missing data through adaptive designs
What Can Advocates Do?
What Will You Do?
What’s Can Advocates to Do?

- Become knowledgeable about innovative designs & Bayesian statistics
- Ask researchers if they have considered using alternative designs
- Advocate for more funding of statistical research, education, and software tools
- Work with the FDA to accelerate release of their recently announced guidance documents