Adaptive Clinical Trials: A Necessary Step toward Personalized Medicine

Donald A. Berry
dberry@mdanderson.org

THE UNIVERSITY OF TEXAS
MD Anderson Cancer Center
Making Cancer History®

Berry Consultants
Statistical Innovation
Financial Disclosure

- Part owner Berry Consultants, LLC
- Designs adaptive trials for
  - Pharmaceutical companies
  - Medical device companies
  - NIH-sponsored cooperative groups
Janet Woodcock, 
Director CDER FDA

“Improved utilization of adaptive and Bayesian methods” could help resolve low success rate of and expense of phase 3 clinical trials

“uncovered a consensus that the two most important areas for improving medical product development are biomarker development and streamlining clinical trials.”

http://www.fda.gov/ScienceResearch/SpecialTopics/CriticalPathInitiative/default.htm
For example, in 2010, the Biomarkers Consortium—a public-private partnership that includes the NIH, the FDA, patient groups, and pharmaceutical and biotech—initiated a groundbreaking trial in breast cancer to predict drug responsiveness based on the presence or absence of genetic and biological markers, … I-SPY 2 (ClinicalTrials.gov NCT01042379).
"Our current approach [to trials] is horribly inefficient, and we need to do something better," says Roger Lewis, an emergency medicine physician at Harbor-University of California, Los Angeles, Medical Center. Lewis helps advise a company called Berry Consultants …

FDA is trying to move forward nevertheless, in part by linking up with more flush agencies. Last week, in conjunction with the National Institutes of Health (NIH), it announced four sizable grants, totaling $9.4 million, in regulatory science. (FDA contributed just under $1 million and NIH gave the rest.) They include support for a heart-lung system that can test potential drugs and an effort to dramatically streamline clinical trials.

"Our current approach [to trials] is horribly inefficient, and we need to do something better," says Roger Lewis, an emergency medicine physician at Harbor-University of California, Los Angeles, Medical Center. Lewis helps advise a company called Berry Consultants founded by Donald Berry, a biostatistician at M.D. Anderson Cancer Center in Houston, Texas. He and Berry, along with emergency medicine physician William Barsan at the
Lewis and Berry, along with emergency medicine physician William Barsan at the University of Michigan, will be studying whether "adaptive" trial designs that incorporate new information in midcourse can answer medical questions. They also want to learn what concerns researchers might have about this approach.
Current use of Bayesian adaptive designs

- MDACC (> 300 trials)
- Device companies (> 25 PMAs)*
- Drug companies (Most of top 40; many biotechs)**

*http://www.fda.gov/MedicalDevicesDeviceRegulationandGuidance/GuidanceDocuments/ucm071072.htm
Example: Troxacitabine in AML* (endpoint: CR by day 50)

- Idarubicin
- Ara-C
- Trox
- Idarubicin
- Trox
- Ara-C

Adaptive randomization to learn, while effectively treating patients

n = 25

* Giles JCO 2003
Adaptive Randomization

- Assign with higher probability to better performing therapies
- TI dropped after 24th patient
- Trial stopped after 34 patients
Summary of AML trial results

CR by 50 days:

<table>
<thead>
<tr>
<th>Group</th>
<th>Count/Total</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>IA</td>
<td>10/18</td>
<td>56%</td>
</tr>
<tr>
<td>TA</td>
<td>3/11</td>
<td>27%</td>
</tr>
<tr>
<td>TI</td>
<td>0/5</td>
<td>0%</td>
</tr>
</tbody>
</table>
Some areas of application of Bayesian adaptive drug trials

- Migraine
- Oncology
- Rh Arthritis
- Lupus
- Diabetes
- Obesity
- Stroke
- Gastroparesis

- Spinal Cord Injury
- HIV
- Hepatitis C
- Pre-term labor
- Constipation
- Overactive bladder
- Libido
- Alzheimer’s
Simulations Usually Required

- To find operating characteristics:
  - Type I error rate
  - Power
  - Sample size distribution
- Prospective design essential
- Longitudinal modeling
- Many scenarios
- Accrual rate matters
"Our current approach [to trials] is horribly inefficient, and we need to do something better," says Roger Lewis, an emergency medicine physician at Harbor-University of California, Los Angeles, Medical Center. Lewis helps advise a company called Berry Consultants ...
Lewis and Berry, along with emergency medicine physician William Barsan at the University of Michigan, will be studying whether "adaptive" trial designs that incorporate new information in midcourse can answer medical questions. They also want to learn what concerns researchers might have about this approach.
Bayesian adaptive trials

- Stopping early (or late)
  - Efficacy
  - Futility
- Dose finding (& dose dropping)
- Seamless phases
- Population finding
- Adaptive randomization
- Ramping up accrual
Why?

- Smaller trials (usually!)
- More accurate conclusions
- Can focus on better treatment of patients in trials
Three Recent Examples of Smaller Sample Size Using

- Bayesian predictive probabilities
- Longitudinal modeling
A Bayesian statistical design was used with a range in sample size from 600 to 1800 patients.
From “Methods”

“These interim analyses were not the standard type in which the trial results are announced when a boundary is crossed. Rather, the decision to discontinue enrollment was based on a prediction that future follow-up was likely to give a meaningful answer.”
Comparison of Antiarrhythmic Drug Therapy and Radiofrequency Catheter Ablation in Patients With Paroxysmal Atrial Fibrillation
A Randomized Controlled Trial

David J. Wilber, MD
Carlo Pappone, MD, PhD
Petr Neuzil, MD
Angelo De Paola, MD
Frank Marchlinski, MD
Andrea Natale, MD
Laurent Macle, MD
Emile G. Daoud, MD
Hugh Calkins, MD
Burr Hall, MD
Vivek Reddy, MD

Context Antiarrhythmic drugs are commonly used for prevention of recurrent atrial fibrillation (AF) despite inconsistent efficacy and frequent adverse effects. Catheter ablation has been proposed as an alternative treatment for paroxysmal AF.

Objective To determine the efficacy of catheter ablation compared with antiarrhythmic drug therapy (ADT) in treating symptomatic paroxysmal AF.

Design, Setting, and Participants A prospective, multicenter, randomized (2:1), unblinded, Bayesian-designed study conducted at 19 hospitals of 167 patients who did not respond to at least 1 antiarrhythmic drug and who experienced at least 3 AF episodes within 6 months before randomization. Enrollment occurred between October 25, 2004, and October 11, 2007, with the last follow-up on January 19, 2009.

Intervention Catheter ablation (n=106) or ADT (n=61), with assessment for effectiveness in a comparable 9-month follow-up period.

Main Outcome Measures Time to protocol-defined treatment failure. The pro-

Design, Setting, and Participants A prospective, multicenter, randomized (2:1), unblinded, Bayesian-designed study conducted at 19 hospitals of 167 patients who in the catheter ablation group remained free from protocol-defined treatment failure compared with 16% of patients treated with ADT. The hazard ratio of catheter ablation to ADT was 0.30 (95% confidence interval, 0.19-0.47; P<.001). Major 30-day treatment-related adverse events occurred in 5 of 57 patients (8.8%) treated with ADT and 5 of 103 patients (4.9%) treated with catheter ablation. Mean quality of life scores improved significantly in patients treated by catheter ablation compared with ADT at 3 months; improvement was maintained during the course of the study.

Conclusion Among patients with paroxysmal AF who had not responded to at least
Time to Chronic Failure by Randomization Group (Updated)

<table>
<thead>
<tr>
<th>TCool</th>
<th>10</th>
<th>69</th>
<th>69</th>
<th>66</th>
<th>63</th>
<th>62</th>
<th>61</th>
<th>54</th>
<th>52</th>
<th>37</th>
<th>15</th>
<th>3</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAD</td>
<td>56</td>
<td>39</td>
<td>29</td>
<td>19</td>
<td>16</td>
<td>13</td>
<td>11</td>
<td>10</td>
<td>7</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Adaptive Design-Key Features

Uses accumulating data to decide on how to modify aspects of the study without undermining the validity and integrity of the trial.

Response-Adaptive Randomization
Uses interim data and Bayesian techniques to modify the treatment allocation probabilities in favor of more informative treatment arms.
- Maximize information about dose-response relationship
- Increases the probability of success in Phase 3 by bringing the correct dose forward into Phase 3

Predefined Decision Rules
Frequent interim analyses allows determination of futility or success
- Allows Ph2a and Ph2b to be combined in a single trial
- Succeed efficiently and fail efficiently

Trial simulations
Define operating characteristics (false + rate, statistical power, and sample size distributions)

Robert Lenz, San Diego, Oct 5, 2009
Response-Adaptive Randomization

1. New Patient
2. Randomization to placebo or optimal dose
3. Ongoing Pts Data
4. Longitudinal model predicts final outcome
5. Estimation of Dose-Response
6. Decision
7. Futility
8. Sufficient information obtained
9. Continue until max. sample size reached
10. Find optimal dose

Robert Lenz, San Diego, Oct 5, 2009
Area of the D-R relationship we want greatest information about

Robert Lenz, San Diego, Oct 5, 2009
Goal is to Find the $ED_{90}$ and MED

$ED_{90}$: smallest dose that produces 90% of the maximal efficacy over the doses studied

MED: minimally effective dose; operationally defined as the smallest dose that produces at least 1.75 point improvement over placebo in ADAS-Cog

Robert Lenz, San Diego, Oct 5, 2009
Adaptive Randomization

- The probability of allocating the subsequent subject to a particular dose is proportional to the reduction in the variance of ED$_{90}$ achieved when a subject is allocated to that dose.

- The allocation ratio will be altered in such a way as to maximize the information (minimize the variance) about the MED and ED$_{90}$.

- Randomization probabilities are updated every 2 weeks.

Robert Lenz, San Diego, Oct 5, 2009
Normal Dynamic Linear Model

• Normal Dynamic Linear Model is used to describe the dose-response relationship.

• This is a very versatile and flexible model that allows for non-monotonic response functions.

• Allows borrowing strength from neighboring doses (don’t treat each dose as entirely independent), thereby narrowing the probability distribution around the dose.

• Thus, increasing the number of doses increases information about the entire dose-response curve, not just a single dose.
Simulations: Sensitivity Analyses

- The primary simulation is based on the best guesses of accrual rate and correlation between early (week 4 and 8) and final time point (week 12) for the ADAS-Cog scores.
- Maximum sample size is 400
- When considering all the simulated D-R curves, very favorable operating characteristics
  - False positive rate less than 5%
  - False negative rate < 20%
- Average sample size across D-R relationships is 319

Robert Lenz, San Diego, Oct 5, 2009
Model estimates of ACAScog (every two weeks)
May 14, 2009: Stop for futility

N = 322
Completers = 238
Same as previous, but dropping incomplete patients
Sample sizes
May 14, 2009: Stop for futility
Final
Adaptive Phase II/III Trial
Randomize

Agent 1
Agent 2
Combo
Control

Randomize

Agent 2
Combo
Control

Randomize

Agent 2
Control

Start
Drop Agent 1
Drop Combo
End

Interim analyses to drop arms and possibly stop the trial

Confirmatory stage

IoM Report on Cancer Cooperative Groups 2010
Example from Critical Path Initiative

- Type II diabetes
- Seamless Phase II/III: Dose finding plus confirmation
- Sample size 200 - 1566
- Active comparator & placebo
- Primary endpoint: Clinical Utility Index (12 months)
Some Details

- Phase II: 7 doses experimental drug plus placebo plus active control
- Phase III
  - 1 or 2 doses experimental drug
  - Sample size via predictive power considering available Phase II data
  - Adaptive transition: Bayesian predictive probs
- Both phases driven by CUI
Clinical Utility Index

- Dose-response modeling
- Longitudinal modeling
Adaptive Biomarker Trial
Randomize throughout

Experimental therapy

Control

Start

Interim analyses to drop non-responding subsets

Confirmatory stage
Responding population

Whole population

Biomarker subsets
Savings possible in sample size when using biomarkers ...
Relapse-free survival in CALGB 9344; n = 2376

- **ER-** and **HER2-**
  - No T: HR = 0.89 (0.79-0.99), N = 681, P = 0.027
  - T: No significant difference

- **ER+** and **HER2-**
  - No T: HR = 1.01 (0.92-1.10), N = 1226, P = 0.95
  - T: No significant difference

- **ER-** and **HER2+**
  - No T: HR = 0.73 (0.59-0.89), N = 192, P = 0.0018

- **ER+** and **HER2+**
  - No T: HR = 0.77 (0.65-0.92), N = 277, P = 0.028

Berry et al. SABCS 2009
New trial design

Uses genetic profiles to highlight 'biomarker' differences among patients and to match drugs to patients with biomarkers that predict a benefit.
I-SPY2: The Cartoon
(Press conference* slides)

*<http://ispy2.org>
Standard Phase 2 Cancer Drug Trials

Population of patients

Outcome: Tumor shrinkage?

Experimental arm

Population of patients

Outcome: Longer time disease free

RANDOMIZE

Experimental arm

Standard therapy
Standard Phase 2 Cancer Drug Trials

Population of patients

Population of patients

Outcome: Tumor shrinkage?

Outcome: Long time disease free

Consequence:
34% success rate of Phase 3 Trials

Standard therapy
I-SPY2 TRIAL

Outcome: Complete response at surgery

Population of patients

Adaptively Randomize

Experimental arm 1
Experimental arm 2
Experimental arm 3
Experimental arm 4
Experimental arm 5
Standard therapy
I-SPY2 TRIAL

Population of patients

Randomly

Arm 2 graduates to small focused Phase 3 trial

Outcome: Complete response at surgery

Experimental arm 1
Experimental arm 2
Experimental arm 3
Experimental arm 4
Experimental arm 5
Standard therapy

Outcome: Complete response at surgery
I-SPY2 TRIAL

Population of patients

RANDOMLY

Arm 3 drops for futility

Experimental arm 1
Experimental arm 3
Experimental arm 4
Experimental arm 5
Standard therapy

Outcome: Complete response at surgery

Outcome: Complete response at surgery
I-SPY2 TRIAL

Population of patients

Randomly

Arm 5 graduates to small focused Phase 3 trial

Experimental arm 1

Experimental arm 4

Experimental arm 5

Standard therapy

Outcome: Complete response at surgery
I-SPY2 TRIAL

Goal: Greater than 85% success rate in Phase 3, with focus on patients who benefit added to the mix.
A New Rx for Medicine

Fed up with slow drug trials, cancer patients and doctors are testing a fast track to personalized treatments.

By RON WINSLOW

PERSONALIZED MEDICINE | How redesigning a clinical trial can speed drug development

Traditional clinical trial
Takes essentially all patients with a disease being studied and is typically intended to eliminate differences in patient characteristics that could bias measures of drug effectiveness.

New trial design
Uses genetic profiles to highlight ‘biomarker’ differences among patients and to match drugs to patients with biomarkers that predict a benefit.

PHASE II
Randomized or non-randomized trial: In a randomized trial, about 60 patients are put in two groups: One receives the experimental drug and the other serves as a control group. In a non-randomized trial, about 40 patients receive the experimental drug.

PHASE III
If a drug graduates to phase III, it typically takes 3,000 patients and about three years to determine if it is safe and effective enough for approval.

HISTORIC SUCCESS RATE
30 TO 40%

PHASE III
Researchers expect that drugs graduating from phase I to phase III can be tested with 300 patients selected according to genetic profiles found to respond to the drug in phase I. It is hoped that this will shorten the time to approval.

PROBABILITY OF SUCCESS
85%

Source: Donald Berry, M.D. Anderson Cancer Center
I-SPY2 Adaptive Design Process

- PI: Laura Esserman, UCSF
- Sponsored by FNIH: NCI, FDA, industry, academia,
- Coordinated with FDA (CDER, CBER & CDRH) from inception
- Current status: 20 centers, ~50 patients, experimental drugs so far: neratinib, ABT888, AMG386
I-SPY 2 Effects

- Match drugs (& combos) with biomarker signatures
- Graduate drug/biomarker pairs to smaller (n < 300), more focused, more successful Phase 3
- Descendents of I-SPY 2 in melanoma, colorectal cancer, Alzheimer’s, acute heart failure, …
Software?
SAS ADAPT?
BeST: Solution for Smarter Clinical Trials
Also, Selected Short Subject ...
Motivated by dinner with Joel Tsevat and John Perentesis ...

Designing a clinical trial is making a decision
Decision analysis as a medical research paradigm

Why do we carry out clinical trials of experimental products?

(1) To see if they are safe and effective.
(2) To see if the results are unusual assuming product is ineffective.
(3) To deliver good medicine to patients.

Does your answer matter?
Standard Approach to Choosing Sample Size

- Consider time to event. Want
  - 25% reduction in hazards

Where are the questions about the disease? the population? Is this CABG or a rare pediatric cancer?

- Median for control: 8 months
- Follow-up after accrual: 12 months

- Answer: n = 650
We need a new paradigm!
Choosing sample size using decision analysis

- Goal: Effective overall treatment of patients, both
  - those who come after the trial and
  - those in the trial
- Example formalizing this goal: Maximize expected number of successes over all patients
“Many respondents viewed the main societal purpose of clinical trials as benefiting the patient rather than creating generalizable knowledge to advance medical science. The fact that many trials involve very small proportions of patients in trials, conflicts with established principles [from Belmont Report] of research ethics.”

We must rethink the “established principles of research ethics”
“More ethical” approach: Maximize overall benefit

- What is “overall”?  
- All patients who will be treated with therapies assessed in trial  
- Call it N, “patient horizon”  
- Enough to know magnitude of N: 100? 1000? 1,000,000?
Goal: maximize expected number of successes in N
Either one- or two-armed trial
Suppose \( n = 1000 \) is right for one trial & \( N = 1,000,000 \)
Then for other N’s use \( n = \)

<table>
<thead>
<tr>
<th>Optimal sample size for one trial and first of two trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
</tr>
<tr>
<td>One trial</td>
</tr>
<tr>
<td>One/two</td>
</tr>
</tbody>
</table>

Ratio of sample sizes within row is general
Ratio across rows applies for particular prior distribution
Figure 33-12  Three different prior distributions for $r$, rate of success. Under distribution A, $r$ is equally likely to be any value between 0 and 1. The density in B is proportional to $r$, which means, for example, that $r$ greater than 0.5 is three times as probable as $r$ less than 0.5. (These two distributions are the same as the two in Figure 33-1, “a.”) Under distribution C, all the probability is concentrated on $r = 0.5$ and so in this case the arm’s effectiveness is assumed to be known.
Optimal allocations in a two-armed trial
<table>
<thead>
<tr>
<th>N</th>
<th>Prior distributions</th>
<th>Succ prop</th>
<th>Prior distributions</th>
<th>Succ prop</th>
<th>Prior distributions</th>
<th>Succ prop</th>
<th>Prior distributions</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>6</td>
<td>5</td>
<td>.63</td>
<td>4</td>
<td>8</td>
<td>.71</td>
<td>9</td>
</tr>
<tr>
<td>1000</td>
<td>21</td>
<td>20</td>
<td>.65</td>
<td>16</td>
<td>30</td>
<td>.74</td>
<td>29</td>
</tr>
<tr>
<td>10,000</td>
<td>70</td>
<td>69</td>
<td>.66</td>
<td>56</td>
<td>98</td>
<td>.75</td>
<td>99</td>
</tr>
<tr>
<td>Large N</td>
<td>$\sqrt{\frac{N}{2}}$</td>
<td>$\sqrt{\frac{N}{2}}$</td>
<td>$\frac{2}{3}$</td>
<td>$\sqrt{\frac{N}{3}}$</td>
<td>$\sqrt{N}$</td>
<td>$\frac{3}{4}$</td>
<td>$\sqrt{N}$</td>
</tr>
</tbody>
</table>
Conclusions

- Standard clinical research paradigm doesn’t apply to rare populations
- Decision analysis addresses right questions, hypothesis testing does not
- Bayesian adaptive approach balances experimentation and treating trial participants effectively
- Bayesian attitude uses all info
- “All info” includes longitudinal info and borrowing adult info to children