Financial Disclosure

Donald Berry is co-owner of Berry Consultants, LLC.

Berry Consultants designs adaptive clinical trials for

- Pharmaceutical companies
- Medical device companies
- NIH cooperative groups
“Improved utilization of adaptive and Bayesian methods” could help resolve low success rate of and expense of phase III 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).
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/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071072.htm

Some areas of application of Bayesian adaptive drug trials

- Oncology
- Migraine
- Rheum Arthritis
- Lupus
- Sepsis
- Diabetes
- Obesity
- Stroke
- Acute heart failure
- Spinal Cord Injury
- HIV
- Hepatitis C
- Pre-term labor
- Constipation
- Micturition
- Alzheimer’s
- Parkinson’s
- Pandemic flu (H1N1)
Types of adaptive trials

- Stopping early (or late)
  - Efficacy
  - Futility
- Dose finding (& dose dropping)
- Seamless phases
- Population finding
- Adaptive randomization
- Selecting end points
Why?

- Smaller trials (usually!)
- More accurate conclusions
- Addresses many questions in one trial
- Better treatment of patients in trials
I-SPY2

http://www.ispy2.org

http://clinicaltrials.gov/ct2/show/NCT01042379?term=I-SPY2&rank=1
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 trials: about 60 patients are put in two groups. One group receives the drug and the other serves as a control group. 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.

PHASE III
Researchers expect that drugs graduating from phase 2 to phase 3 can be tested with 300 patients selected according to genetic profiles found to respond to the drug in phase II. It is hoped that this will shorten the time to approval.

PROBABILITY OF SUCCESS
85%

Source: Donald Berry, M.D., Anderson Cancer Center
THE SATURDAY ESSAY  |  OCTOBER 2, 2010

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 WINSWOLD

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 II
Patients are placed in groups based on genetic profiles and are randomly assigned to either standard therapy or one of five different drugs plus standard care.

Early results increase chances that patients entering the trial later will be assigned to a drug showing benefit against tumors with their genetic profile.

PHASE III
Researchers expect that drugs graduating from I-Spy 2 to phase III can be tested with 300 patients selected according to genetic profiles found to respond to the drug in phase II. It is hoped that this will shorten the time to approval.

PROBABILITY OF SUCCESS
85%

Note: In all clinical trials, phase I consists of testing on human subjects to determine toxicity levels.

Graphic by Maryanne Manary/WSJ
I-SPY 2 Adaptive Trial Design

Paclitaxel* +
Investigational Agent A
(12 weekly cycles)

Paclitaxel*
(12 weekly cycles)

Paclitaxel* +
Investigational Agent B
(12 weekly cycles)

AC
(4 cycles)

AC
(4 cycles)

AC
(4 cycles)

MRI
Biopsy
Blood Draw

MRI
Blood Draw

MRI
Biopsy
Blood Draw

MRI
Blood Draw

Tissue

* HER2 positive participants will also receive trastuzumab. An investigational agent may be used instead of trastuzumab.
I-SPY2: The Cartoon

*<http://ispy2.org>
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

Outcome: Complete response at surgery

Population of patients

Arm 2 graduates to small focused Phase 3 trial

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

Population of patients

Arm 3 drops for futility

Outcome: Complete response at surgery

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

Complete response at surgery for Arm 3 drops for futility

Outcome: Complete response at surgery

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

Outcome: Complete response at surgery

Arm 5 graduates to small focused Phase 3 trial

Population of patients

Randomly

Experimental arm 1
Experimental arm 4
Experimental arm 5
Standard therapy

D. Berry, San Antonio, 22 Mar 2013
I-SPY2 TRIAL

Population of patients

Outcome:
Complete response at surgery

Arm 6 is added to the mix

Experimental arm 1
Experimental arm 4
Experimental arm 6
Standard therapy

ADAPTIVELY
RANDOMIZED

D. Berry, San Antonio, 22 Mar 2013
I-SPY-like TRIAL for Combinations

Outcome: pathCR or PFS or OS

Population of patients

Adaptively

A
B
C
D
A + SOC
B + SOC
C + SOC
D + SOC
C + D + SOC
SOC
I-SPY-like TRIAL for Combinations

Substudy: Adaptively randomized factorial

Population of patients

 Outcome: pathCR or PFS or OS

A + SOC
B + SOC
C + SOC
D + SOC
C + D + SOC
SOC
Goal: Greater than 85% success rate in Phase III, with focus on patients who benefit
Substudy: Adaptively randomized factorial
I-SPY2 Adaptive Process

- PI: Laura Esserman, UCSF
- Considers 10 biomarker signatures
- Never-ending screening process
- Sponsored by FNIH: NCI, FDA, industry, academia
- Coordinated with FDA (CDER, CBER, & CDRH)—Regulatory pathway
- Status: 19 centers, 400 pts randomized, first 7 exp drugs: neratinib, ABT888, AMG386, AMG479, MK2206, pertuzumab, pertuzumab+T-DM1
Experimental Drugs

- Sample size for each drug, 20 to 120
- Up to 8 exp drugs at a time
- Covariate modeling (across subtypes)
- Longitudinal modeling using MRI
- Assign in proportion to current (Bayesian) probability drug >> control, by subtype
Dropping, Graduating Drugs

- For each possible biomarker signature \( S \), calculate probability drug \( \gg \) control in \( S \).
- If Bayesian predicted probability 300-pt Phase III success < 10% for all \( S \), drop drug.
- If > 85% for some \( S \) then drug graduates.
- At graduation, we provide predictive probability Phase III success for each \( S \), including \( S \) on drug’s diploma.
I-SPY2 and I-SPY3 Effects

- Match drugs with biomarker signatures
- Savings from common control
- Better therapies move thru faster
- Successful drug/biomarker pairs graduate to small (n < 300), focused, more successful Phase 3 based on Bayesian predictive probabilities
- Offspring of I-SPY 2: melanoma, GBM, Lymphoma, Alzheimer’s, HIV, acute heart failure, SARI/H1N1, …
## Patient Strata Are Fixed

### pCR by Subtype in I-SPY1

<table>
<thead>
<tr>
<th></th>
<th>MP−</th>
<th>MP+</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR+</td>
<td>HR−</td>
</tr>
<tr>
<td>HER2+</td>
<td>0.47</td>
<td>0.67</td>
</tr>
<tr>
<td>HER2−</td>
<td>0.25</td>
<td>0.43</td>
</tr>
</tbody>
</table>

**MP:** MammaPrint High+ vs High-
**HR+:** Hormone Receptor+: Either ER+ or PgR+ (No trastuzumab)
## Estimated Prevalences from I-SPY1

<table>
<thead>
<tr>
<th></th>
<th>MP-</th>
<th></th>
<th>MP+</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR+</td>
<td>HR-</td>
<td>HR+</td>
<td>HR-</td>
</tr>
<tr>
<td>HER2+</td>
<td>16%</td>
<td>7%</td>
<td>4%</td>
<td>10%</td>
</tr>
<tr>
<td>HER2-</td>
<td>23%</td>
<td>6%</td>
<td>6%</td>
<td>28%</td>
</tr>
</tbody>
</table>
Biomarker signatures

- Graduate drugs/signatures from trial:
  - Based on effectiveness
  - Based on prevalence

- Biomarker signatures (2^8-1 combinations of subtypes): S₁, S₂, ..., S₂⁵⁵

- But restrict to (10) marketable signatures:

<table>
<thead>
<tr>
<th></th>
<th>MP-</th>
<th>MP+</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR+</td>
<td>HR-</td>
</tr>
<tr>
<td>HER2+</td>
<td>16%</td>
<td>7%</td>
</tr>
<tr>
<td>HER2-</td>
<td>23%</td>
<td>6%</td>
</tr>
</tbody>
</table>
pCR is itself a surrogate marker!
Guidance for Industry
Pathologic Complete Response in Neoadjuvant Treatment of High-Risk Early-Stage Breast Cancer: Use as an Endpoint to Support Accelerated Approval

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)

May 2012
Clinical/Medical
Designs for I-SPY 3:
A Biomarker-Signature
Neoadjuvant Breast Cancer
Trial to Address Confirmation
of I-SPY 2 Graduates for
Both pCR and EFS
Association of pCR with EFS in Triple Negative Subtype

**Graph:**
- **Event-Free Survival Probability**
- **Months since Randomization**
- **HR=0.24, P* < 0.001 (CI: 0.18 to 0.33)**
- **pCR (n = 389)**
- **no pCR (n = 768)**

**Legend:**
- Green line: pCR
- Red line: no pCR

**Notes:**
- pCR=ypT0/is ypN0
- (Courtesy of Patricia Cortazar)
- * Nominal p-value
Association of pCR with EFS in Triple Negative Subtype

**Triple Negative**

![Graph showing Event-Free Survival Probability over Months since Randomization for pCR and no pCR](image)

- **HR=0.24, P* < 0.001**
  - CI: 0.18 to 0.33
- **pCR (n = 389)**
- **no pCR (n = 768)**

(Courtesy of Patricia Cortazar)

pCR=ypT0/is ypN0

* Nominal p-value
Distributions of Hazards (CTNeoBC), Assuming Exponential
Smoothed Version of Cortazar

[Graph showing event-free survival over months for pCR and No pCR]
EFS for Various pCR Rates
EFS for pCR Rates 70% Vs 35%

Event-free Survival

Months

0 10 20 30 40 50 60 70 80 90

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

No pCR

Control

Exp

pCR

HR ~ 0.62
EFS for pCR Rates 55% Vs 35%

- Control: EFS for pCR Rates 55% Vs 35%
- Exp: EFS for pCR Rates 55% Vs 35%
- No pCR: EFS for pCR Rates 55% Vs 35%

HR ~ 0.75
For Designs Considered

- Accelerated approval if superiority on pCR rate
- pCR analysis when all patients have surgery
- Single final pCR test: Chi-square
- Full approval if superiority on EFS
- 3 years minimum follow-up for EFS
- Single log-rank test for EFS
- Type I error rate controlled ≤ 2.5%
- 20/mo accrual
Power Via Simulations: Fixed Design
(Same for Group-Sequential)
When designing a trial, assuming a particular “Clinically Significance Difference” is statistically convenient ... and scientifically naïve!

Point Estimates Are Killers!
Typical(?) Uncertainty in pCR Rate after Phase II

7/20 (35%) Controls
22/40 (55%) Experimental  28/40 (70%) Experimental
Joint Density of pCR Rates: 7/20 Controls Vs 22/40 Exp
Contour Plot: 7/20 Control Vs 22/40 Exp

Odds ratio = 2

Equal pCR rates
Fixed 1200

<table>
<thead>
<tr>
<th>Power</th>
<th>Predictive Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.884</td>
<td>1.000</td>
</tr>
<tr>
<td>0.695</td>
<td>0.856</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Both</th>
<th>pCR</th>
<th>EFS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power</td>
<td>0.884</td>
<td>1.000</td>
</tr>
<tr>
<td>PPower</td>
<td>0.695</td>
<td>0.856</td>
</tr>
</tbody>
</table>

In the graph, the pCR rate and mean pCR rate are plotted against power and predictive power. The graph shows the relationship between these variables and how they affect the outcomes of different treatments or conditions.
Better Information about pCR Rates
—More Relevant to I-SPY 3—
Would Greatly Improve I-SPY 3’s Efficiency
Goldilocks Design

• Maximum sample size $N (= 1200)$

• When $300^{th}$ pt has surgery, find predictive probabilities of both pCR and EFS statistically significant based on pCR data (only!) from I-SPY 3 (plus CTNeoBC)
  - If PP < 5% when $N$ then stop now for futility
  - If PP > 90% with current $n$ then stop accrual (final $n$ greater by ~120)

• Else continue to next 100 surgeries; repeat above until $n = N$

• In all cases, pCR analysis after 6 mos, EFS analysis after 3 yrs
Goldilocks Vs Fixed Vs One IA

<table>
<thead>
<tr>
<th></th>
<th>Both</th>
<th>pCR</th>
<th>EFS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed</td>
<td>0.0036</td>
<td>0.025</td>
<td>0.025</td>
</tr>
<tr>
<td>Gold</td>
<td>0.0028</td>
<td>0.013</td>
<td>0.024</td>
</tr>
<tr>
<td>Fixed</td>
<td>0.884</td>
<td>1.000</td>
<td>0.884</td>
</tr>
<tr>
<td>Gold</td>
<td>0.883</td>
<td>0.999</td>
<td>0.883</td>
</tr>
</tbody>
</table>
Mean Sample Size

![Graph showing mean sample size across different pCR rates for Fixed1200, One1200, and Gold1200.](image)

- **Fixed1200**
- **One1200**
- **Gold1200**

The graph illustrates how the mean sample size changes with varying pCR rates for the indicated conditions.
Goldilocks Vs Fixed Vs One IA
Accounting for Unknown pCR Rates

Predictive Power

Power/Predictive Power vs pCR rate/Mean pCR rate
Mean Sample Size
Based on Predictive Probabilities

Mean N

Mean pCR rate

Fixed1200
One1200
Gold1200
MRD a surrogate marker in leukemia?
Example: MRD/Relapse in ALL

Holowiecki BJH 2008
Example: MRD/Relapse in ALL

- MRD < 0.1%
- MRD > 0.1%

Holowiecki BJH 2008
Tumor Agnostic ("Basket Trial") (Trial ongoing in pharma)

- Targeted drug, develop simultaneously across organ-specific cancers, restricting to tumors that express target
- Population sizes small means sample sizes necessarily small
- Endpoint: tumor response
The Approaching Wall

- Ever finer grid of biomarker categories—within 10 years every cancer patient will have an orphan disease

- How to develop drugs in this setting?
Approaches/Opportunities

- Difficult inferential/regulatory problem
- Formal, reproducible approach desired
- Hierarchical modeling across tumor types is a possible approach
- Adaptive stopping within histology
- There may be others to be considered
Hierarchical modeling; Bayes borrowing assumptions

Population of response rates within 10 tumor types:

Response rates $p_i$ have a distribution, one that is imperfectly known, even after observing the $R_i/N_i$. 

Observations
Hierarchical modeling; Bayes borrowing assumptions

Population of response rates within 10 tumor types:

R_1/N_1 gives info about p_1 which gives info about population of p’s which gives info about p_2, say. Hence “borrowing.” [Charles Stein & history ...]
Hierarchical modeling; Assumptions & prior

- Distribution of response rates $p_i$ is unknown—it itself has a probability distribution.

- Expectations regarding $p$’s can differ by tumor type.

- Prior distribution ("hyperprior") of heterogeneity $\sigma$ in population of $p$’s is important in determining borrowing.
Learn about heterogeneity parameter $\sigma$ from trial results

$\sigma$ large:

$\sigma$ small:
Many Possibilities …

σ moderate:

σ moderate:

Better: clustering …
Example Analysis
Data = (4,8,5,3,7,5,5,6,3) N=10 per group

95% credible interval for each ORR with NO borrowing. Circle is the posterior mean. Data is a “noisy” version of ORR. Estimates are farther apart than actual ORR values. Makes sense to “shrink”.
Example Analysis

Data = (4,8,5,3,7,5,5,6,3)  N=10 per group

Red lines show pooled analysis.
Shrinkage is extreme.
Want something more moderate.
Example Analysis

Data = (4, 8, 5, 3, 7, 5, 5, 6, 3)  N=10 per group

Blue lines show hierarchical model approach.
Shrinkage based on removing binomial error, but leaving variation among the individual histologies (assumed to be a normal distribution)
Histologies allowed to vary. This shrinks, but far from pooling.
Hierarchical modeling shows shrinkage. Goal of clustering model is to isolate histologies such as 3 and 4 by borrowing only from similar histologies.
Example Analysis
Data = (3,2,7,6,1,1,1,0,1)  N=10 per group

Green shows clustered hierarchical models. Isolates histologies 3 and 4 (shrinking half as much). Histologies 1, 2, 5, 6, 7, 8, 9 regress to each other.