Adaptive Bayesian Trials

Donald A. Berry
dberry@mdanderson.org
I am a co-owner and part-time consultant with Berry Consultants, LLC, a company that designs adaptive clinical trials for pharmaceutical companies, medical device companies, NIH cooperative groups, international consortia.
A Historical Perspective on Clinical Trials Innovation and Leadership
Where Have the Academics Gone?

David L. DeMets, PhD
Robert M. Califf, MD

The randomized controlled trial (RCT), the gold standard for evaluating the balance of risk and benefit in medical therapies, first emerged as a key clinical research tool in the mid-20th century thanks to visionary leadership of agencies such as the US National Institutes of Health (NIH), the UK Medical Research Council, and academic research institutions. Since then, clinical trials activity has shifted from the NIH and academia into the purviews of the medical products industry and regulatory authorities. Recent emphasis on evidence-based medicine, patient-centered outcomes research, and learning and accountable health care systems underscores the fact that most clinical trials fail to provide the evidence needed to inform medical decision making. However, the serious implications of this deficit are largely absent from public dis-

When fundamental trials methodologies were being developed at the NIH in the 1960s, an NIH-commissioned task force delineated recommendations for organizing and conducting RCTs. One significant early example is the Coronary Drug Project, a joint effort among NIH sponsors, an academic coordinating center, and a steering committee of academic leaders. In the 1970s and 1980s, the NIH often convened academic leaders to identify knowledge gaps and prioritize and conduct specific trials as funding permitted.

During the 1960s, there was scant statistical literature examining clinical trials methodologies. Researchers learned by doing trials, noting successes and failures, and iterating to advance the field. In a series of discussions in the 1970s, ideas were debated and solutions to immediate problems were proposed. Throughout the 1970s and 1980s, NIH and academic biostatisticians developed many methods now in routine use, including sample size estimation, interim data monitoring, and repeated measure methods for analysis.

At the outset of this era, few large randomized clinical
IS IT TIME TO RETIRE THE RANDOMIZED TRIAL?

This symposium will explore whether or not the evolution of observational research methods for making fair-comparisons on the one hand and the mushrooming ethical, logistical, and financial barriers to performing RCTs on the other, have finally tipped the balance. Drawing speakers and panelists from international, continental and local centres of excellence, we will present the issues, arguments, and counter-arguments in open plenary sessions, and provide plenty of times and places for their informal exploration and debate.

Sessions on randomized clinical trials, observational studies and where the twain shall meet:

- Past successes
- Strengths and weaknesses
- Potential achievements
- Problems and concerns
- New perspectives and solutions

TARGET AUDIENCE:

SPEAKERS:
Donald Berry
John Concato
PJ Devereaux
Susan Elbonberg
Dean Follmann
Steven Goodman
David Henry
Ralph Horwitz
Malcolm Macleod
Klim McPherson
Donald Redelmeier
David Sackett
Holger Schünemann
Stan Shapiro
Robert Temple
Peter Tugwell
Sean Tunis
Jan Vandenbroucke
Michael Walsh
Salim Yusuf

University of Texas
Yale University
McMaster University
University of Pennsylvania
US National Institute of Allergy and Infectious Disease
Stanford University
Institute for Clinical Evaluative Sciences (ICES)
GlaxoSmithKline & Yale University (Emeritus)
University of Edinburgh
University of Oxford
Sunnybrook Health Sciences Centre
McMaster University (Emeritus)
McMaster University
McGill University
U.S. Food and Drug Administration (USFDA)
University of Ottawa
Center for Medical Technology Policy (CMTP)
Universiteit Leiden
McMaster University
McMaster University

PLANNING COMMITTEE:
Holger Schünemann
McMaster University, Committee Chair

John Fletcher
Canadian Medical Association Journal

David Sackett
McMaster University (Emeriti)

Peter Tugwell
University of Ottawa
Can Big Data Tell Us What Clinical Trials Don’t?

When a helicopter rushed a 13-year-old girl showing symptoms suggestive of kidney failure to Stanford’s Packard Children’s Hospital, Jennifer Frankovich was the rheumatologist on call. She and a team of other doctors quickly diagnosed lupus, an autoimmune disease. But as they hurried to treat the girl, Frankovich thought that something about the patient’s particular combination of lupus symptoms — kidney problems, inflamed pancreas and blood vessels — rang a bell. In the past, she’d seen lupus patients with these symptoms develop life-threatening blood clots. Her colleagues in other specialties didn’t think there was cause to give

Yet despite the pitfalls, developing a “learning health system” — one that can incorporate lessons from its own activities in real time — remains tantalizing to researchers. Stefan Thurner, a professor of complexity studies at the Medical University of Vienna, and his researcher, Peter Klimek, are working with a database of millions of people’s health-insurance claims, building networks of relationships among diseases. As they fill in the network with known connections and new ones mined from the data, Thurner and Klimek hope to be able to predict the health of individuals or of a population over time. On the clinical side, Longhurst has been advocating for a button in electronic medical-record software that would allow doctors to run automated searches for patients like theirs when no other sources of information are available.

With time, and with some crucial refinements, this kind of medicine may eventually become mainstream. Frankovich recalls a conversation with an older colleague. “She told me, ‘Research this decade benefits the next decade,’” Frankovich says. “That was how it was. But I feel like it doesn’t have to be that way anymore.”
OUTLINE

- Intro to Bayes
  - Predictive probabilities
  - Longitudinal modeling
- Adaptive trials
- I-SPY 2 (and I-SPY 3)
- Basket trials
Bayes Rule

Rule of inverse probabilities:

\[ P(H|\text{data}) \propto P(\text{data}|H)*P(H) \]

Familiar application: positive predictive value of diagnostic test:

\[ P(\text{dis}|+) \propto P(+|\text{dis})*P(\text{dis}) \]
Bayes Rule

Rule of inverse probabilities:

\[ P(H|\text{data}) \propto P(\text{data}|H)*P(H) \]
Bayesian Approach

- Formalism for learning under uncertainty
- Conditions on everything known
- Probabilities for all unknowns: hypotheses, future data
- Inherently synthetic
- Hypothesis test: Probability of treatment effect
- Interval estimation: Interval that contains parameter with 95% probability
Advantages of Bayes

- Naturally adaptive — on-line learning (e.g., predictive probability)
- Uses early by-patient information (via longitudinal modeling)
- Can use historical information (e.g., via hierarchical modeling)
Bayesian Updating

- Paired observations, T vs C
- $r = \text{success rate} = P(T \text{ wins pair})$
- $H_0: r = 1/2$
- Data: SSFSS FSSSF
Prob: 8/12

Prob: 4/12
Example (cont’d)

● Full Data:

SSFSS  FSSSF
SFSSS  SS

● 13 S's and 4 F's

● Updated density

\[ \propto r^{13} (1-r)^4 \]
Current (posterior) for success rate $r$

$r^{13}(1 - r)^4$
Given 13 S’s of 17, and assuming uniform prior on $r$.

Suppose double the sample size—next 17 observations:
Predictive probabilities vs. best fitting binomial

**Predictive,** $p \sim \text{beta}(14,5)$

- 88% probability of statistical significance

**Binomial,** $p = 14/19$

- 96% probability of statistical significance
Bayesian Clinical Trial with Smaller Sample Size

- 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.”
Predicting trial results

- Simulate
- Model/parameter uncertainty
- Incorporate info (Bayesian-wise) on various outcomes
- Model relationships among early and late endpoints
- Consider alternative designs
Power considerations

![Diagram showing power considerations with H0 and H1]

- Traditional power
- Underlying treatment difference

Power

H0

H1
True (predictive) power

Underlying treatment difference

H₀  H₁

Power

Traditional power
Predictive power
Prior density
NXY-059 for Acute Ischemic Stroke

Kennedy R. Lees, M.D., Justin A. Zivin, M.D., Tim Ashwood, Ph.D.,
Antonio Davalos, M.D., Stephen M. Davis, M.D., Hans-Christoph Diener, M.D.,
James Grotta, M.D., Patrick Lyden, M.D., Ashfaq Shuaib, M.D.,
Hans-Göran Härdemark, M.D., and Warren W. Wasiewski, M.D.,
for the Stroke—Acute Ischemic NXY Treatment (SAINT I) Trial Investigators*
RESULTS
Among the 1699 subjects included in the efficacy analysis, NXY-059 significantly improved the overall distribution of scores on the modified Rankin scale, as compared with placebo ($P=0.038$ by the Cochran–Mantel–Haenszel test). The common odds ratio for improvement across all categories of the scale was 1.20 (95 percent confidence interval, 1.01 to 1.42).
<table>
<thead>
<tr>
<th>Outcome Variable</th>
<th>Placebo Group</th>
<th>NXY-059 Group</th>
<th>Difference between NXY-059 and Placebo†</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% or score (95% CI)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modifed Rankin scale score (primary end point)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of patients</td>
<td>849</td>
<td>850</td>
<td>4.4</td>
<td></td>
</tr>
<tr>
<td>Score — no. (%)</td>
<td>93 (11.0)</td>
<td>131 (15.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>170 (20.0)</td>
<td>153 (18.0)</td>
<td>-2.0</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>99 (11.7)</td>
<td>97 (11.4)</td>
<td>-0.3</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>108 (12.7)</td>
<td>121 (14.2)</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>175 (20.6)</td>
<td>144 (16.9)</td>
<td>-3.7</td>
<td></td>
</tr>
<tr>
<td>5 (or death)</td>
<td>204 (24.0)</td>
<td>204 (24.0)</td>
<td>0</td>
<td>0.038</td>
</tr>
<tr>
<td>Change from baseline in total NIHSS score (coprimary outcome)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of patients</td>
<td>851</td>
<td>851</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Score — LSM ±SE</td>
<td>-1.7±0.5</td>
<td>-1.8±0.5</td>
<td>-0.1 (-1.4 to 1.1)</td>
<td>0.86</td>
</tr>
<tr>
<td>Barthel index (dichotomized analysis)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of patients</td>
<td>848</td>
<td>850</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Score, ≥95 — no. (%)</td>
<td>346 (40.8)</td>
<td>368 (43.3)</td>
<td>2.5</td>
<td>0.14</td>
</tr>
<tr>
<td>Stroke Impact Scale</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of patients</td>
<td>676</td>
<td>669</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Score — LSM ±SE</td>
<td>63.4±1.1</td>
<td>66.2±1.1</td>
<td>2.8 (-0.3 to 5.9)</td>
<td>0.08</td>
</tr>
<tr>
<td>EuroQoL EQ-5D (weighted index)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of patients</td>
<td>816</td>
<td>819</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Score — LSM ±SE</td>
<td>0.43±0.013</td>
<td>0.47±0.013</td>
<td>0.04 (-0.0 to 0.07)</td>
<td>0.06</td>
</tr>
<tr>
<td>EuroQoL EQ-5D (VAS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of patients</td>
<td>671</td>
<td>670</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Score — LSM ±SE</td>
<td>62.0±0.9</td>
<td>64.5±0.9</td>
<td>2.5 (-0.1 to 5.0)</td>
<td>0.05</td>
</tr>
</tbody>
</table>

P-value is NOT probability of null hypothesis
SAINT II:

- N = 3200 (up from 1700)
- Power 80% for Odds Ratio 1.20
Naïve predictive power of SAINT II = 60%

My probability that SAINT II would be positive: 10%
Press release, Oct 26, 2006

“Results from the SAINT II (Stroke Acute Ischemic NXY-059 Treatment) trial: … NXY-059 did not meet its primary outcome of a statistically significant reduction in stroke-related disability, as assessed by the modified Rankin Scale (mRS) (p=0.33, odds ratio 0.94) compared to placebo.

“The company plans no further development of NXY-059 in acute ischemic stroke.”
The bitterest pill

Drug companies lose hundreds of millions of dollars when human clinical trials fail. Helen Pearson examines why some procedures could help avoid such disappointments.

Cape Town in October has many attractions: balmy temperatures, enticing beaches, simple wines and the indomitable presence of Table Mountain. But for the many neurologists attending the Joint World Congress on Stroke there last month, it was a grim place to be.

The neurologists learned that a highly promising stroke drug called NXY-059 had dismally failed in a phase III clinical trial. The results, announced by AstraZeneca on 26 October, were particularly disappointing because they contradicted those of a previous phase III trial, which had shown that the same drug could help patients by quenching the free radicals that damage neurons after a stroke.

AstraZeneca's shares fell by 7.5% to $61.38 on the day of the announcement. Shares in Renovis, the San Francisco biotechnology company that licensed the drug to AstraZeneca, plunged by more than 75%. Neurologists at the meeting were just as deflated. "We were all pretty down and asking what the hell's going on here," recalls NXY-059, which had contradicted those of a previous phase III trial, which had shown that the same drug could help patients by quenching the free radicals that damage neurons after a stroke.

AstraZeneca's shares fell by 7.5% to $61.38 on the day of the announcement. Shares in Renovis, the San Francisco biotechnology company that licensed the drug to AstraZeneca, plunged by more than 75%. Neurologists at the meeting were just as deflated. "We were all pretty down and asking what the hell's going on here," recalls NXY-059, which had contradicted those of a previous phase III trial, which had shown that the same drug could help patients by quenching the free radicals that damage neurons after a stroke.

But of more than major clinical trials decades after stroke has proven effective because the human brain area or underlying disease animal models. Even so, hope 059, which had tested in animals, was borne out after the patients (see K. 354, 588–600; 2001). But there were trials. It showed it better than a placebo using one scale to measure improvement in a patient's disability — but it didn't significantly improve symptoms, using a second, separate, measure of its impact. This made at least some specialists question whether the drug would really improve patients' quality of life. AstraZeneca was "pushing the data trials in humans fail to reach the approval stage. Consultancy McKinsey found that 42% of them were unsuccessful. In half of these failures, the drugs were no more effective than a placebo; 30% raised safety concerns, and the remaining 20% were found to be no safer or more effective.
Most of the effect is regression to the mean
Irreproducible Research and Multiplicities

Two PCAST panels on Jan 31, 2014:
http://www.tvworldwide.com/events/pcast/140131/
OUTLINE

- Intro to Bayes
  - Predictive probabilities
  - Longitudinal modeling
- Adaptive trials
- I-SPY 2 (and I-SPY 3)
- Basket trials
Bayesian adaptive trials

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

- Smaller trials (usually!)
- More accurate conclusions
- Kill duds early, and soundly
- Can focus on better treatment of patients in trials
Example: Troxacitabine in AML*
(endpoint: CR by day 50)

Randomize

Idarubicin
Ara-C
n = 25

Trox
Idarubicin
n = 25

Trox
Ara-C
n = 25

Standard design

* Giles JCO 2003
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 in trial

* 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:

IA  10/18 = 56%
TA  3/11 = 27%
TI  0/5  =  0%
Adaptive Clinical Trials
A Partial Remedy for the Therapeutic Misconception?

William J. Meurer, MD, MS; Roger J. Lewis, MD, PhD; Donald A. Berry, PhD


There is a common “therapeutic misconception” among patients considering participation in clinical trials. Some trial participants and family members believe that the goal of a clinical trial is to improve their outcomes—a misperception often reinforced by media advertising of clinical research. Clinical trials have primarily scientific aims and rarely attempt to collectively improve the outcomes of their participants. The overarching goal of most clinical trials is to evaluate the effect of a treatment on disease outcomes. Comparisons are usually made with placebo for conditions having no established treatments and with standard care for conditions having effective treatments. Any benefit to an individual trial participant is a chance effect of randomization and the true, but unknown, relative effects of the treatments. Available evidence is conflicting regarding whether patients receive some benefit from simply participating in a clinical trial. Thus, even though serving as a research participant is essentially an altruistic activity, many
Adaptive Trial with Longitudinal Modeling

- Type II diabetes
- Seamless Phase II/III: Dose finding via adaptive randomization then confirmation
- Active comparator & placebo
- Primary endpoint: Clinical Utility Index (12 months)
Some Details

● Phase II: 7 doses experimental drug, adaptively randomized

● Seamless switch to Phase III
  ■ 1 or 2 doses experimental drug
  ■ Sample size via predictive power considering available Phase II data
  ■ Adaptive transition: Bayesian pred probs

● Longitudinal modeling critical

● Both phases driven by CUI
Clinical Utility Index

- Dose-response modeling
- Longitudinal modeling is critical
Bayesian adaptive trials

- Stopping early (or late)
  - Efficacy
  - Futility
- Dose finding (& dose dropping)
- Seamless phases
- Population finding
- Adaptive randomization
- Modifying accrual rate

Primary efficacy endpoints, as measured by reduction in hemoglobin A1c (HbA1c) at the 1.5mg dose, were met in three studies (AWARD-1, AWARD-3 and AWARD-5). Having met the primary endpoints, superiority for HbA1c lowering was examined, and both doses of dulaglutide (0.75mg and 1.5mg) demonstrated statistically superior reduction in HbA1c from baseline compared to: exenatide twice-daily injection at 26 weeks (AWARD-1); metformin at 26 weeks (AWARD-3); and sitagliptin at 52 weeks (AWARD-5).
FDA News Release

FDA approves Trulicity to treat type 2 diabetes

For Immediate Release  September 18, 2014

The U.S. Food and Drug Administration today approved Trulicity (dulaglutide), a once-weekly subcutaneous injection to improve glycemic control (blood sugar levels), along with diet and exercise, in adults with type 2 diabetes.

Type 2 diabetes affects about 26 million people and accounts for more than 90 percent of diabetes cases diagnosed in the United States. Over time, high blood sugar levels can increase the risk for serious complications, including heart disease, blindness, and nerve and kidney damage.

"Type 2 diabetes is a serious chronic condition that causes blood glucose levels to get too high, increasing the risk for serious complications," said Dr. Robert E. Goldberg, M.D., Acting Director, Office of Rare Diseases and Ophthalmology, Center for Drug Evaluation and Research. "FDA's approval of Trulicity provides another option for people with type 2 diabetes and helps further expand our therapeutic toolkit to address this serious condition."

Trulicity is the only once-weekly glucagon-like peptide-1 (GLP-1) receptor agonist approved for once-weekly injection in adults with type 2 diabetes. GLP-1 receptor agonists, also known as incretin mimetics, lower blood sugar levels by stimulating the release of insulin from the pancreas and decreasing the amount of glucose produced by the liver.
Adaptive Platform Trials …
When the I-SPY 2 trial launched in 2010, oncologists heralded it as the future of cancer research. Five pharmaceutical companies put aside their differences to participate in the landmark phase 2 breast cancer trial, which adaptively and efficiently randomized patients to one of seven experimental therapies. Now, even as I-SPY 2 propels its first two drugs into phase 3 trials, researchers in other areas of medicine are catching on to the benefits of this collaborative approach. On 11 December, Europe's Innovative Medicines Initiative (IMI) announced a €53 million call for proposals for a similarly designed trial in Alzheimer's disease. Already, at least 12 drug companies are keen to participate.
OUTLINE

- Intro to Bayes
  - Predictive probabilities
  - Longitudinal modeling
- Adaptive trials
- I-SPY 2 (and I-SPY 3)
- Basket trials
Prototype Platform Trial: I-SPY 2

http://www.ispy2.org

http://clinicaltrials.gov/ct2/show/NCT01042379?term=I-SPY2&rank=1
I-SPY2 Adaptive Process

- Neoadjuvant breast cancer
- Primary endpoint: pCR (Role of MRI?)
- 10 biomarker signatures
- Never-ending screening process
- First sponsor: FNIH (NCI, FDA, industry)
- Coordinated with FDA (CDER, CBER, & CDRH)—Regulatory pathway
- Status: 20 centers, 680 pts randomized, first 8 exp drugs: neratinib, veliparib, AMG386, AMG479, MK2206, pertuzumab, pertuzumab+T-DM1, ganetespib, plus …

Graduates
Graduation or Futility Met?

- No: Continue Trial
- Yes: Revise Randomization Probabilities within Each Disease Subtype

Calculate Success Prob for Each Signature

Begin Trial with Equal Randomization Probabilities

Accrual Rate Permitting, Add Experimental Arms

Stop Accrual in that Arm
San Antonio Breast Cancer Symposium, Dec 2013

Results of Veliparib/Carboplatin Arm Compared with Control Therapy

The I-SPY 2 Bayesian model provides the probability distributions of pCR rates in each signature
Veliparib/carboplatin graduated with Triple-negative signature

Estimated pCR rate: 26%  Estimated pCR rate: 52%
95% interval: 11% to 40%  95% interval: 35% to 69%
Estimated pCR Rate:
HER2-negative/HR-positive signature

Estimated pCR rate: 14%
95% probability interval: 4% to 27%

Estimated pCR rate: 19%
95% probability interval: 6% to 35%
Neratinib plus standard neoadjuvant therapy for high-risk breast cancer: Efficacy results from the I-SPY 2 TRIAL

John W. Park, Minetta C. Liu, Douglas Yee, Angela DeMichele, Laura van 't Veer, Nola Hylton, Fraser Symmans, Meredith B. Buxton, A. Jo Chien, Amy Wallace, Michelle Melisko, Richard Schwab, Judy Boughey, Debashish Tripathy, Hank Kaplan, Rita Nanda, Stephen Chui, Kathy S. Albain, Stacy Moulder, Anthony Elias, Julie E. Lang, Kirsten Edminston, Donald Northfelt, David Euhus, Qamar Khan, Julia Lyandres, Sarah E. Davis, Christina Yau, Ashish Sanil, Laura J. Esserman, and Donald A. Berry

on behalf of the I-SPY 2 TRIAL Investigators
Neratinib graduated in HR-/HER2+

Pred. prob. 79%

Ctl

Neratinib

pCR rate

33% 56%
Adaptive Platform Effects

- Match drugs with biomarker signatures
- Savings from common control
- Better therapies move thru faster
- Successful drug/biomarker pairs graduate to small, focused, more successful Phase III based on Bayesian predictive probabilities
An advantage of I-SPY 2 platform design

- Indirect comparisons
- Pertuzumab + trastuzumab becomes control in HER2+
<table>
<thead>
<tr>
<th>Characteristics of Modern Platform Trials</th>
<th>I-SPY 2</th>
<th>MICAT</th>
<th>BATTLE</th>
<th>LUNG MAP</th>
<th>UK Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screen markers for all patients</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Master protocol</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Drugs from many companies</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Combination therapies</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Sequential therapies</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Regimens enter &amp; leave trial</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Learn off-target effects</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Pair regimens with biomarkers</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Common control arm</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Adaptive randomization</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Adaptive sample size</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Early “curable” disease</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Registration endpoint</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Seamless phases</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Longitudinal modeling</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>
Triaging is important but no adaptations.

Tissue Submission

Biomarker Profiling*

Biomarker 1: Sub-study 1 (Exp¹ SoC¹)
Biomarker 2: Sub-study 2 (Exp² SoC²)
Biomarker 3: Sub-study 3 (Exp³ SoC³)
Biomarker 4: Sub-study 4 (Exp⁴ SoC⁴)
Not Biomarker 1-4: Non-match Study (NMT SoC⁵)

*Sub-studies assigned based on biomarker results, patients with multiple biomarkers randomly assigned to sub-study.

Exp = Targeted therapy (TT) or TT combinations (TTC), Exp¹-⁴ are different TT/TTC regimens
NMT = non-match study experimental therapy or combinations
SoC = docetaxel or erlotinib, SoC¹-⁵ depends on biomarker and TT/TTC/NMT regimen

5 independent studies with no off-target assessment.
Lung-MAP Sub-Studies for Treatment

Patients with squamous cell lung cancer

Tumor sample analyzed

Sub-Study A
- Sub-Study B
- Sub-Study C
- Sub-Study D
- Sub-Study E

~35%
- Tumor has none of the changes listed here
  - Arm A1
  - Arm A2
  - 50% Chemotherapy
  - Medimmune Anti-PD-L1

~8%
- Tumor DNA has PIK3CA gene mutation
  - Arm B1
  - Arm B2
  - 50% Chemotherapy
  - 50% Pictilisib
  - Genentech GDC-0032

~50%
- Tumor DNA has CCND1, D2, CDK4 gene amplification
  - Arm C1
  - Arm C2
  - 50% Chemotherapy
  - 50% Palbociclib
  - Pfizer Palbociclib

~20%
- Tumor DNA has FGFR gene amplification, mutation or fusion
  - Arm D1
  - Arm D2
  - 50% Chemotherapy
  - 50% AZD 4547
  - AstraZeneca AZD4547

~5%
- Tumor contains high levels of c-Met protein
  - Arm E1
  - Arm E2
  - 50% Erlotinib
  - 50% Rilotumumab mab+ Erlotinib
  - Amgen Rilotumumab

Berry’s estimates
LUNG-MAP design

- Substudies open and close independently (although ...)
- Seamless Phases II and III
- Phase II endpoint is PFS
- Phase II target is 55 events (80 – 150 patients)
- Continue to Phase III if 53% improvement in median PFS: ~1.5 mos
- Agents replaced if miss Phase II goal
- Phase III endpoint is OS, powered for 50% improvement
- Phase III in each substudy ~400 patients
- Crossovers not allowed
- Total duration 2 – 7 years
- Agents studied in biomarker-positives only (although 2012 FDA guidance makes marker-negatives a priority)

Phase II patients count in Phase III — I think.

No longitudinal modeling!

1.5 mos improvement in OS < 20%!
Power for 50% improvement in OS is < 20%

And what about combinations?
OUTLINE

- Intro to Bayes
  - Predictive probabilities
  - Longitudinal modeling
- Adaptive trials
- I-SPY 2 (and I-SPY 3)
- Basket trials
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
Efficient Phase III designs to address both pCR and EFS in a single trial
Powered for both pCR and EFS

- Accelerated approval if superiority on pCR
  ✓ pCR analysis when all patients thru surgery
- Full approval if superiority on EFS
  ✓ 3 years min follow-up for EFS
  ✓ Type I error rate controlled ≤ 2.5%
Cortazar et al.
Triple-negative BC, pCR → EFS

Event-Free Survival Probability

- pCR (n = 389)
- no pCR (n = 768)

HR = 0.24, P < 0.001
(CI: 0.18 to 0.33)

Months since Randomization

69
Updated information about pCR rates and relationships pCR→EFS greatly improves I-SPY 3 efficiency
Power, Goldilocks vs Fixed 1200 Design
Mean Sample Size, Goldilocks vs Fixed 1200

![Graph showing the comparison between Goldilocks and Fixed 1200 in terms of mean sample size vs pCR rate. The graph indicates a peak in the mean sample size at a specific pCR rate, with Goldilocks showing a higher mean sample size compared to Fixed 1200.](image-url)
OUTLINE

- Intro to Bayes
  - Predictive probabilities
  - Longitudinal modeling
- Adaptive trials
- I-SPY 2 (and I-SPY 3)
- Basket trials
Basket Trial Methodology

- Targeted drug, develop simultaneously across organ-specific tumors, restricted to those expressing target
- Sample sizes tiny, borrow but don’t “pool” (formalizes “Gleevec phenomenon”)
Hierarchical modeling/Bayesian borrowing

Population of response rates within tumor types:

Learn about heterogeneity and clustering from trial results

Heterogeneous, little borrowing:

Homogeneous, much borrowing:
Learn about heterogeneity and clustering from trial results

Some heterogeneity, some borrowing:

Outlier #7 stands apart; no borrowing from or to main cluster:
Open dashed circles: raw CBR estimates for 9 histologies
Solid circles: estimates adjusted for borrowing
Area of circle: “equivalent sample size”
Small sample size and further from cluster mean is regressed more
Estimates further from the cluster mean borrow less
Modular phase II study to link targeted therapy to patients with pathway-activated tumors

Signature is an innovative approach to early signal finding

- Rapidly matches patients to therapies that target their molecular abnormality
- Features a rapid trial deployment model and patient-sparing adaptive statistical design

<table>
<thead>
<tr>
<th>PATHWAY</th>
<th>COMPOUND</th>
<th>NCT NUMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td>PI3K, PTEN</td>
<td>Buparlisib (BKM120) 100 mg daily (20 days)</td>
<td>NCT01833169 1</td>
</tr>
<tr>
<td>FGFR, PDGFR, VEGFR, cKIT, FLT3, CSF1R, TrkA, RET</td>
<td>Dowlinib (TKI258) 500 mg daily (5 days on, 2 days off; 28 days)</td>
<td>NCT01831726 2</td>
</tr>
<tr>
<td>RAF, RAS, MEK, NF1</td>
<td>Binimetinib (MEK162) 45 mg bid (28 days)</td>
<td>NCT01886195 3</td>
</tr>
<tr>
<td>BRAF/V600e</td>
<td>Encorafenib (BAY8069) 300 mg daily (28 days)</td>
<td>NCT01081187 4</td>
</tr>
<tr>
<td>PTCH1, SMO</td>
<td>Sonikagib (LDE229) 800 mg daily (28 days)</td>
<td>NCT02002889 5</td>
</tr>
<tr>
<td>ALK, ROS1</td>
<td>Ceritinib (LDK378) 750 mg daily (23 days)</td>
<td>COMING SOON</td>
</tr>
<tr>
<td>CDK4/6, CycIn D1/3, p16</td>
<td>LEE011 600 mg daily (3 weeks on, 1 week off)</td>
<td>COMING SOON</td>
</tr>
<tr>
<td>FGFR</td>
<td>BGJ398 125 mg daily (3 weeks on, 1 week off)</td>
<td>COMING SOON</td>
</tr>
</tbody>
</table>

Patient preidentified with pathway-activated tumor

Study start-up (3 weeks)

Berry Consultants “signature”

Adaptive statistical design

Treatment duration: until unacceptable toxicity, disease progression, death, and/or treatment discontinuation.
Eligibility for NCT018333169

- Heme malignancies and solid tumors except GBM, NSCLC, endometrial, prostate, and breast cancers
- Tumor has activation of PI3K pathway, by CLIA lab
- At least 1 prior treatment for metastatic or locally advanced disease
- Performance status ≤ 1
OUTLINE

● Intro to Bayes
  ■ Predictive probabilities
  ■ Longitudinal modeling

● Adaptive trials

● I-SPY 2 (and I-SPY 3)

● Basket trials