Adaptive Bayesian Trials

Donald A. Berry
dberry@mdanderson.org

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

Financial Disclosure

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.

JAMA, February 16, 2011—Vol 305, No. 7

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,[1] and learning[2] and accountable[3] 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.[4] One significant early example is the Coronary Drug Project,[5] 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.[6] 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

McMaster University
HEALTH SCIENCES

The Second DAVID SACKETT SYMPOSIUM

September 27th & 28th, 2013

Crowne Plaza, Niagara Falls, 5685 Falls Avenue, Niagara Falls, Ontario, L2E 6W7

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 RCT's 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:

Physicians, Specialists, Researchers, Students, Residents, Clinical Trialists, Observationalists, Health Research Methodologists, Health Policy Decision Makers, Diagnosticians, Clinical-practice Guidelines Formulators.

SPEAKERS:

Donald Berry	University of Texas
John Concato	Yale University
PJ Devereaux	McMaster University
Susan Ellenberg	University of Pennsylvania
Dean Follmann	US National Institute of Allergy and Infectious Di
Steven Goodman	Stanford University
David Henry	Institute for Clinical Evaluative Sciences (ICES)
Ralph Horwitz	GlaxoSmithKline & Yale University (Emeritus)
Malcolm Macleod	University of Edinburgh
Klim McPherson	University of Oxford
Donald Redelmeier	Sunnybrook Health Sciences Centre
David Sackett	McMaster University (Emeritus)
Holger Schünemann	McMaster University
Stan Shapiro	McGill University
Robert Temple	U.S. Food and Drug Administration (USFDA)
Peter Tugwell	University of Ottawa
Sean Tunis	Center for Medical Technology Policy (CMTP)
Jan Vandenbroucke	Universiteit Leiden
Michael Walsh	McMaster University
Salim Yusuf	McMaster University

PLANNING COMMITTEE:

Holger Schünemann
McMaster University, Committee Chair

David Sackett
McMaster University (Emeritu

John Fletcher
Canadian Medical Association Journal

Peter Tugwell
University of Ottawa

Can Big Data Tell Us What Clinical Trials Don't?

OCT. 3, 2014

Illustration by Christopher Brand

Eureka

By VERONIQUE GREENWOOD

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:

$$P(H|data) \propto P(data|H)*P(H)$$

Familiar application: positive predictive value of diagnostic test:

$$P(dis|+) \propto P(+|dis)*P(dis)$$

Bayes Rule

Rule of inverse probabilities:

$$P(H|data) \propto P(data|H)*P(H)$$

Bayesian Approach

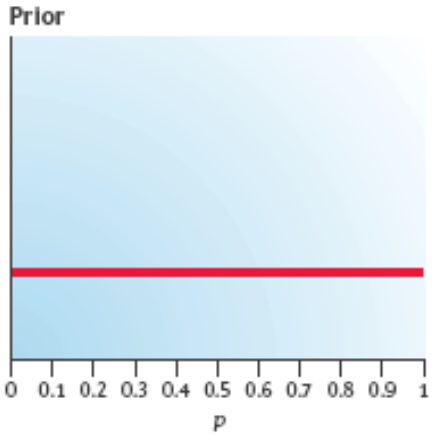
- Formalism for learning under uncertainty
- Conditions on everything known
- Probabilities for all unknowns: hypotheses, future data
- Inherently synthetic
- Hypothesis test: Probability of treatment effect
- Interval estimation: Interval that contains parameter with 95% probability

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 = success rate
  = P(T wins pair)
- H_0: $r = 1/2$
- Data: SSFSS FSSSF



Prob: 8/12

Prob: 4/12

Example (cont'd)

- **Full Data:**

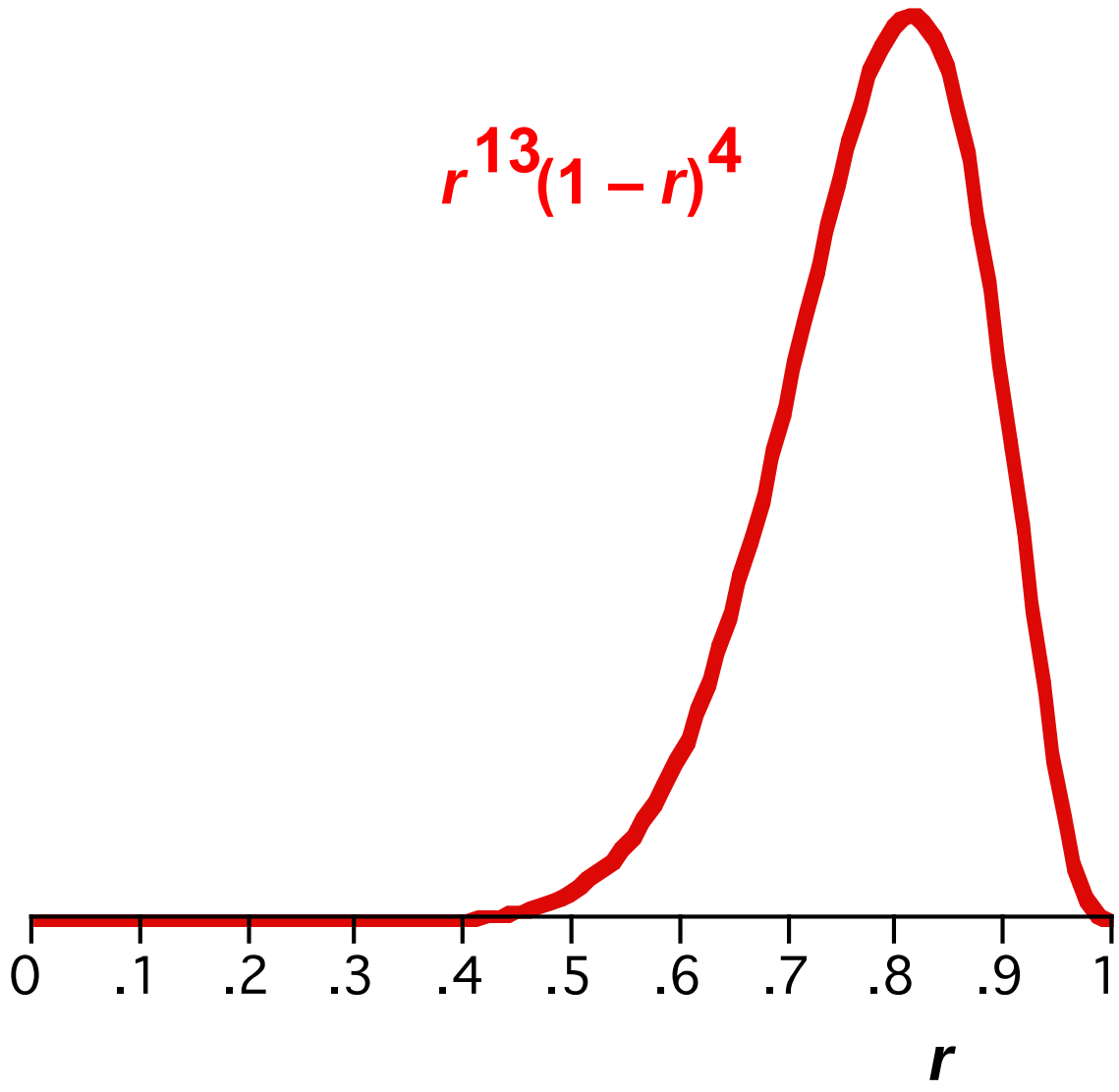
  SSFSS FSSSF
  SFSSS SS

- **13 S's and 4 F's**

- **Updated density**

$$\propto r^{13}(1-r)^4$$

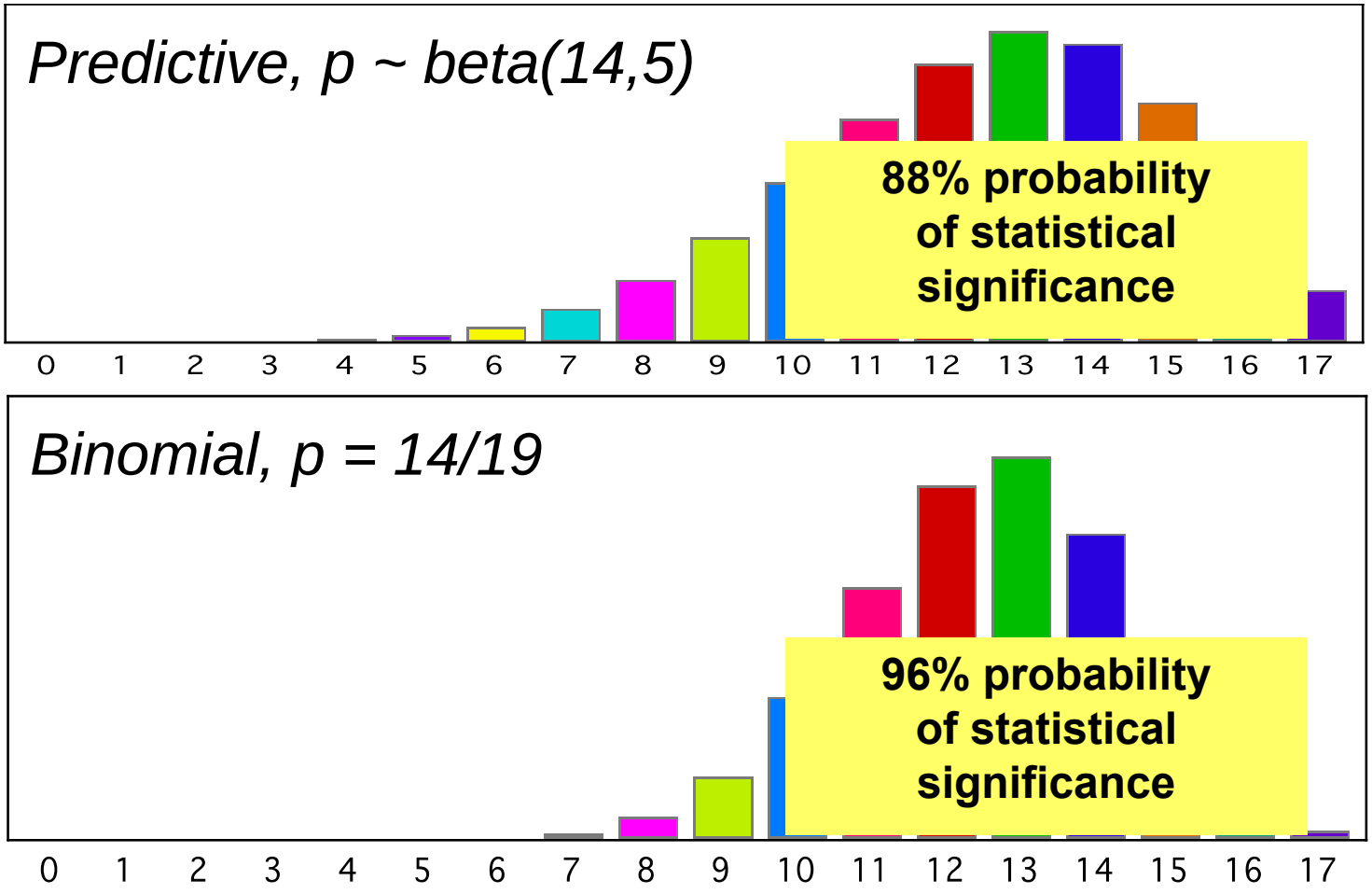
Current (posterior) for success rate *r*



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



Bayesian Clinical Trial with Smaller Sample Size

- Predictive probabilities
- Longitudinal modeling

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 MAY 14, 2009 VOL. 360 NO. 20

Adjuvant Chemotherapy in Older Women with Early-Stage Breast Cancer

Hyman B. Muss, M.D., Donald A. Berry, Ph.D., Constance T. Cirrincione, M.S., Maria Theodoulou, M.D., Ann M. Mauer, M.D., Alice B. Kornblith, Ph.D., Ann H. Partridge, M.D., M.P.H., Lynn G. Dressler, Ph.D., Harvey J. Cohen, M.D., Heather P. Becker, Patricia A. Kartcheske, B.S., Judith D. Wheeler, M.P.H., Edith A. Perez, M.D.,

A Bayesian statistical design was used with a range in sample size from 600 to 1800 patients.

BACKGROUND

Older women with breast cancer are underrepresented in clinical trials, and data on the effects of adjuvant chemotherapy in such patients are scant. We tested for the noninferiority of capecitabine as compared with standard chemotherapy in women with breast cancer who were 65 years of age or older.

METHODS

We randomly assigned patients with stage I, II, IIIA, or IIIB breast cancer to standard chemotherapy (either cyclophosphamide, methotrexate, and fluorouracil or cyclophosphamide plus doxorubicin) or capecitabine. Endocrine therapy was recommended after chemotherapy in patients with hormone-receptor–positive tumors. A Bayesian statistical design was used with a range in sample size from 600 to 1800 patients. The primary end point was relapse-free survival.

From the University of Vermont, Burlington (H.B.M.); the M.D. Anderson Cancer Center, Houston (D.A.B.); the Cancer and Leukemia Group B (CALGB) Statistical Center, Duke University Medical Center (C.T.C., P.A.K.) and Duke University Medical Center (H.J.C., J.D.W., A.A.M.) — both in Durham, NC; Memorial Sloan-Kettering Cancer Center, New York (M.T., L.N., C.A.H.); CALGB, Chicago (A.M.M., H.P.B.); the Dana–Farber Cancer Institute, Boston (A.B.K., A.H.P., H.J.B., E.P.W.); the University of North Carolina, Chapel Hill (L.G.D.); the North Central Cancer Treat-

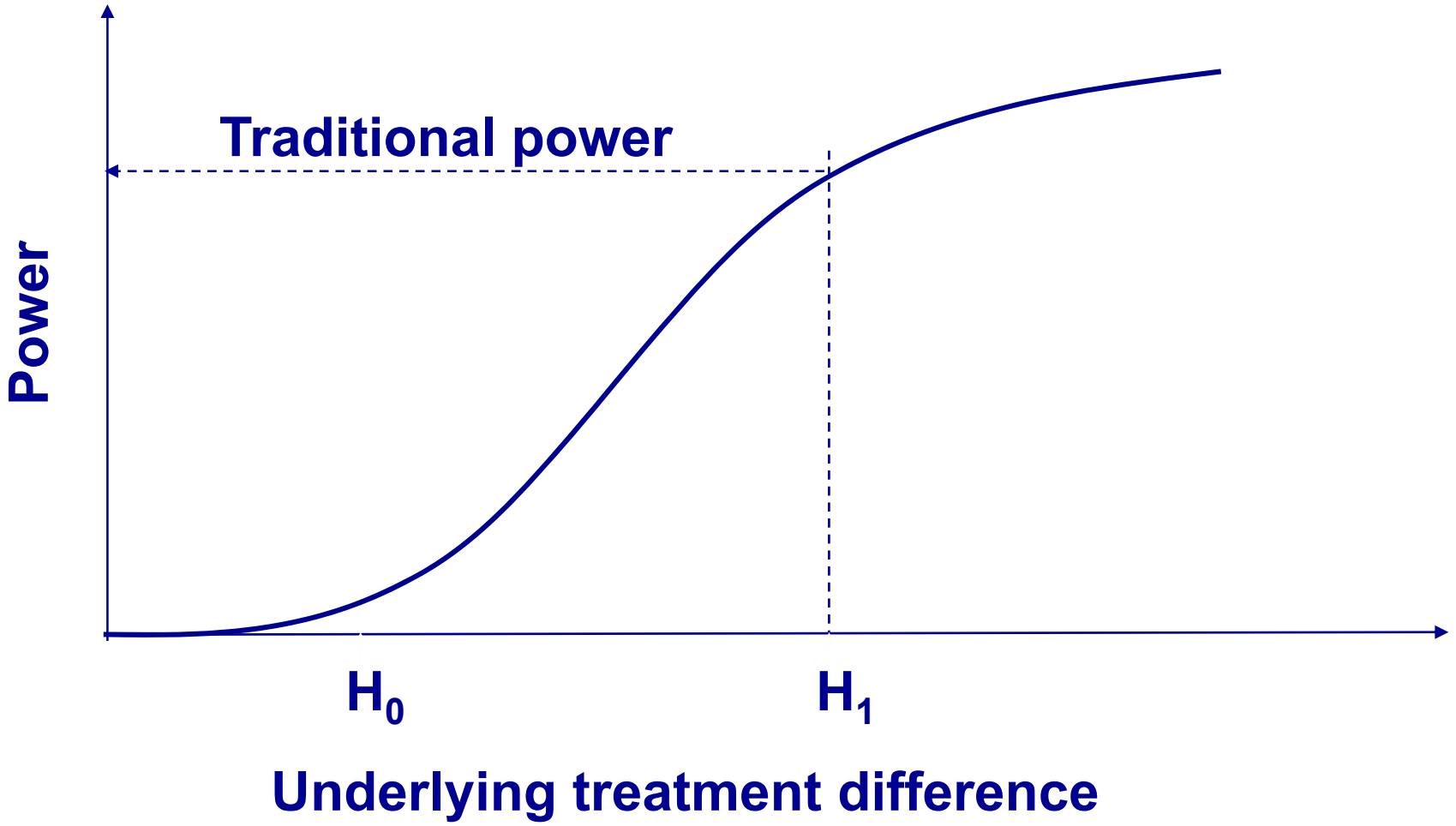
From "Methods"

"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."

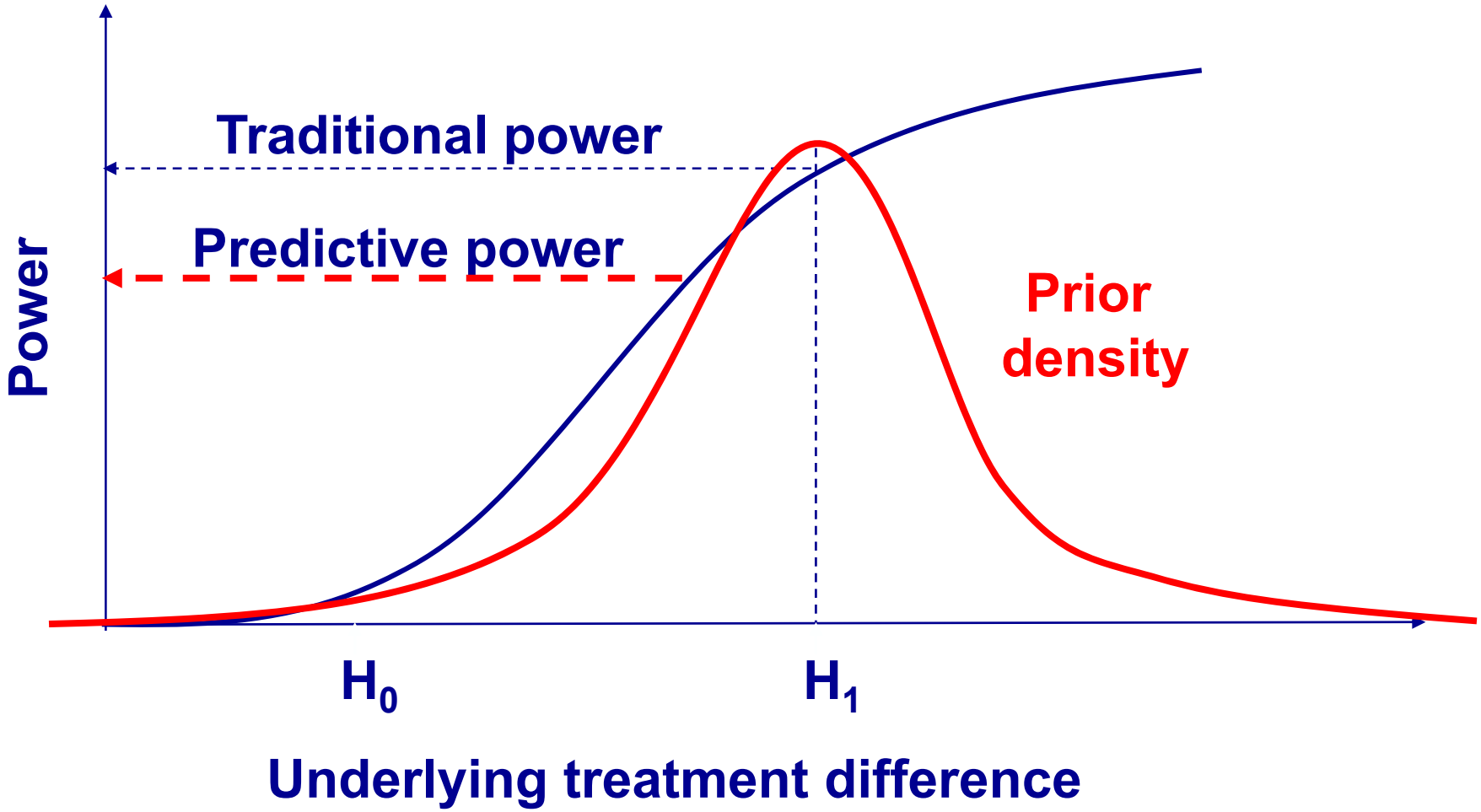
Predicting trial results

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

Power considerations



True (predictive) power



The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

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 Hårdemark, M.D., and Warren W. Wasiewski, M.D., for the Stroke–Acute Ischemic NXY Treatment (SAINT I) Trial Investigators*

N ENGL J MED 354;6 WWW.NEJM.ORG FEBRUARY 9, 2006

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).

?

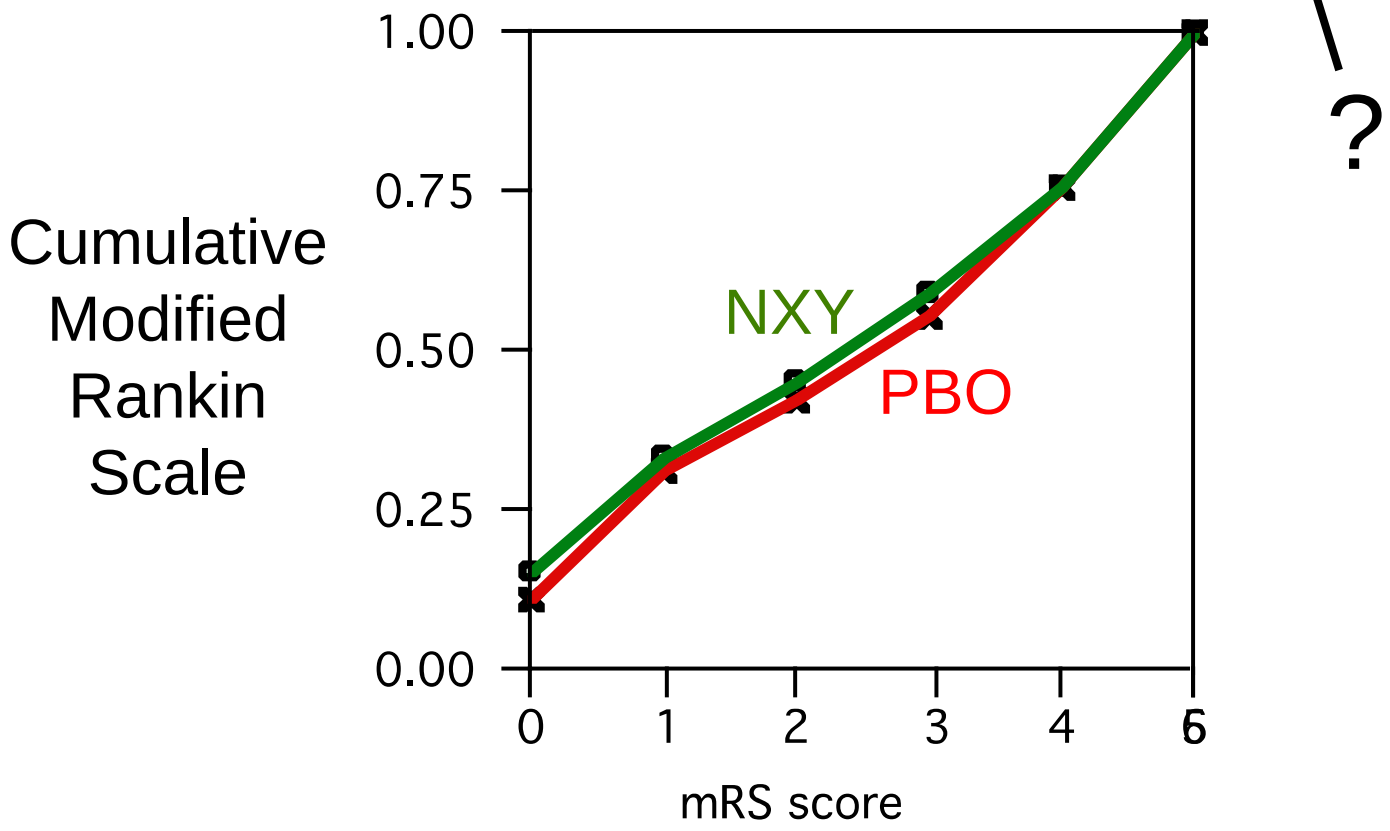


Table 2. Efficacy of the Study Drug at Day 90 or at the Last Rating.*

Outcome Variable	Placebo Group	NXY-059 Group	Difference between NXY-059 and Placebo†	P Value
			% or score (95% CI)	
Modified Rankin scale score (primary end point)				
No. of patients	849	850		
Score — no. (%)				
0	93 (11.0)	131 (15.4)	4.4	
1	170 (20.0)	153 (18.0)	–2.0	
2	99 (11.7)	97 (11.4)	–0.3	
3	108 (12.7)	121 (14.2)	1.5	
4	175 (20.6)	144 (16.9)	–3.7	
5 (or death)	204 (24.0)	204 (24.0)	0	0.038
Change from baseline in total NIHSS score (coprimary outcome)				
No. of patients	851	851		
Score — LSM ±SE	–1.7±0.5	–1.8±0.5	–0.1 (–1.4 to 1.1)	0.86
Barthel index (dichotomized analysis)				
No. of patients	848	850		
Score, ≥95 — no. (%)	346 (40.8)	368 (43.3)	2.5	0.14
Stroke Impact Scale				
No. of patients	676	669		
Score — LSM ±SE	63.4±1.1	66.2±1.1	2.8 (–0.3 to 5.9)	0.08
EuroQoL EQ-5D (weighted index)				
No. of patients	816	819		
Score — LSM ±SE	0.43±0.013	0.47±0.013	0.04 (–0 to 0.07)	0.06
EuroQoL EQ-5D (VAS)				
No. of patients	671	670		
Score — LSM ±SE	62.0±0.9	64.5±0.9	2.5 (–0.1 to 5.0)	0.05

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."

BUSINESS

The bitterest pill

Drug companies lose hundreds of millions of doll
human clinical trials fail. **Helen Pearson** examine
procedures could help avoid such disappointme

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

But of more than
major clinical tr
decades after str
has proven effec
ably because the
human brain are
of underlying dis
animal models.

Even so, hope
059, which had
tests in animals.
borne out after t
patients (see K.
354, 588–600; 20

But there were
trial. It showed t
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

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

PLOS | MEDICINE

Why Most Published Research Findings Are False

John P. A. Ioannidis

Published: August 30, 2005 • DOI: 10.1371/journal.pmed.0020124

Abstract

Summary

There is increas
the number of ot
framework, a res
and lesser presel
and other interes
designs and setti
simply accurate r

THE NEW YORKER

ANNALS OF SCIENCE

THE TRUTH WEARS OFF

Is there something wrong with the scientific method?

BY JONAH LEHRER

DEC

O

in a
start
atyp
on th
bran
been tested on schizophrenics in several large clinical trials, all of which had demonstrated a dramatic decrease in the subjects' psychiatric symptoms. As a result, second-generation antipsychotics had become one of the fastest-growing and most profitable pharmaceutical classes. By 2001, Eli Lilly's Zyprexa was generating more revenue than Prozac. It remains the company's top-selling drug.

The Economist | World politics | Business & finance | Economics | Science & technology | Cult

Unreliable research

Trouble at the lab

Scientists like to think of science as self-correcting. To an alarming degree, it is not

Oct 19th 2013 | From the print edition

Like 20k Tweet 1,984

Many results that are rigorously proved and accepted start shrinking in later studies.

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

NC8817Z

Bayesian adaptive trials

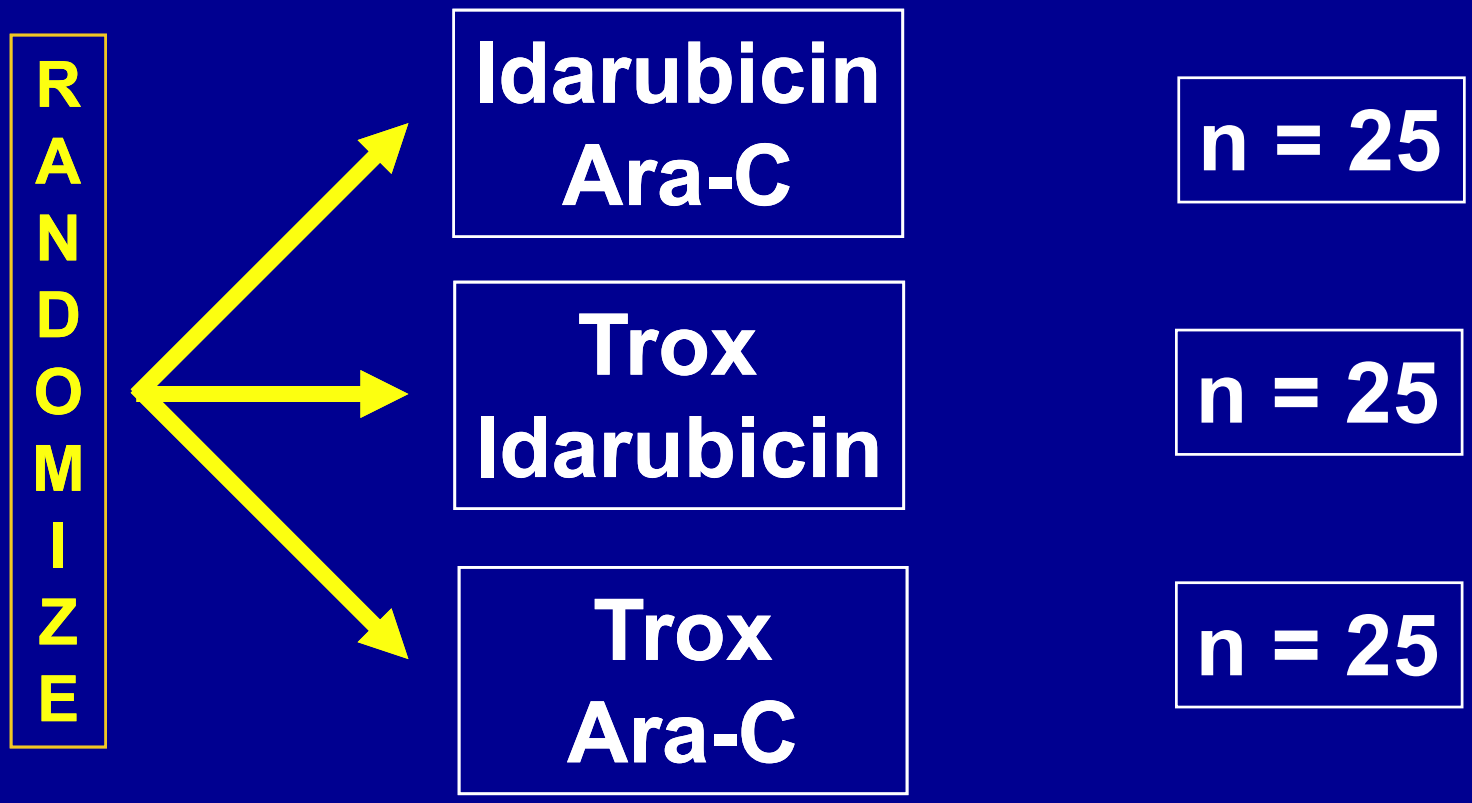
- Stopping early (or late)
  - Efficacy
  - Futility
- Dose finding (& dose dropping)
- Seamless phases
- Population finding
- Adaptive randomization
- Modifying accrual rate

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)

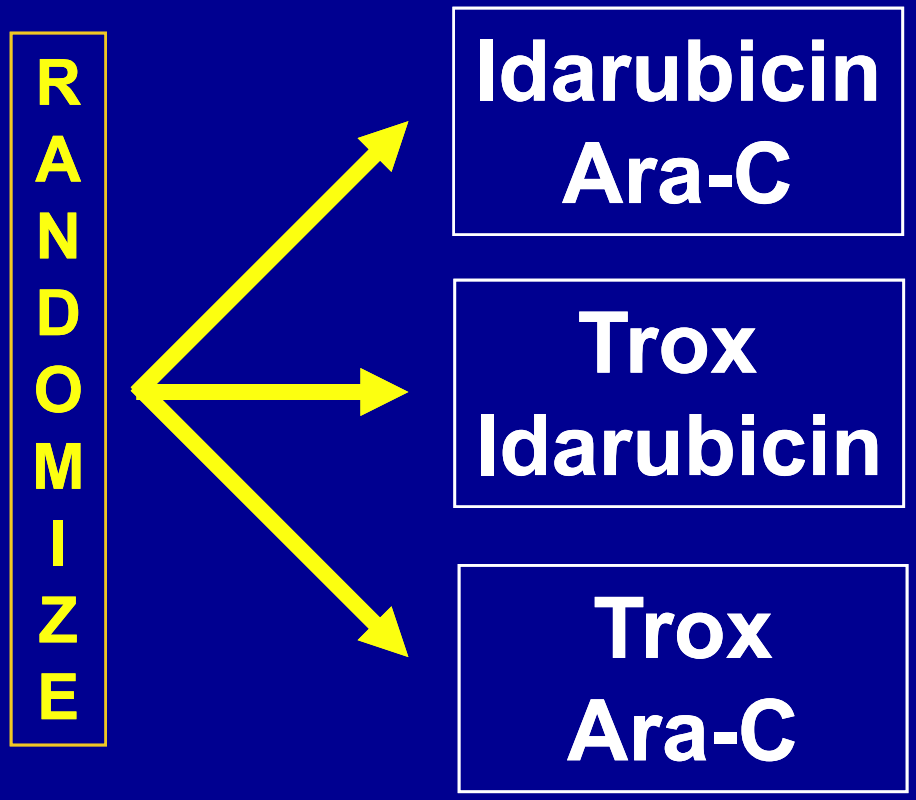
Standard design



* Giles JCO 2003

Example: Troxacitabine in AML* (endpoint: CR by day 50)

Our design



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%

JAMA The Journal of the American Medical Association

June 13, 2012, Vol 307, No. 22 >

Adaptive Clinical Trials
A Partial Remedy for the Therapeutic Misconception?

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

JAMA. 2012;307(22):2377-2378. doi:10.1001/jama.2012.4174.

There is a common "therapeutic misconception" among patients considering participation in clinical trials.[1] 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.[2] 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.[3] 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.[3] Thus, even though serving as a research participant is essentially an altruistic activity, many

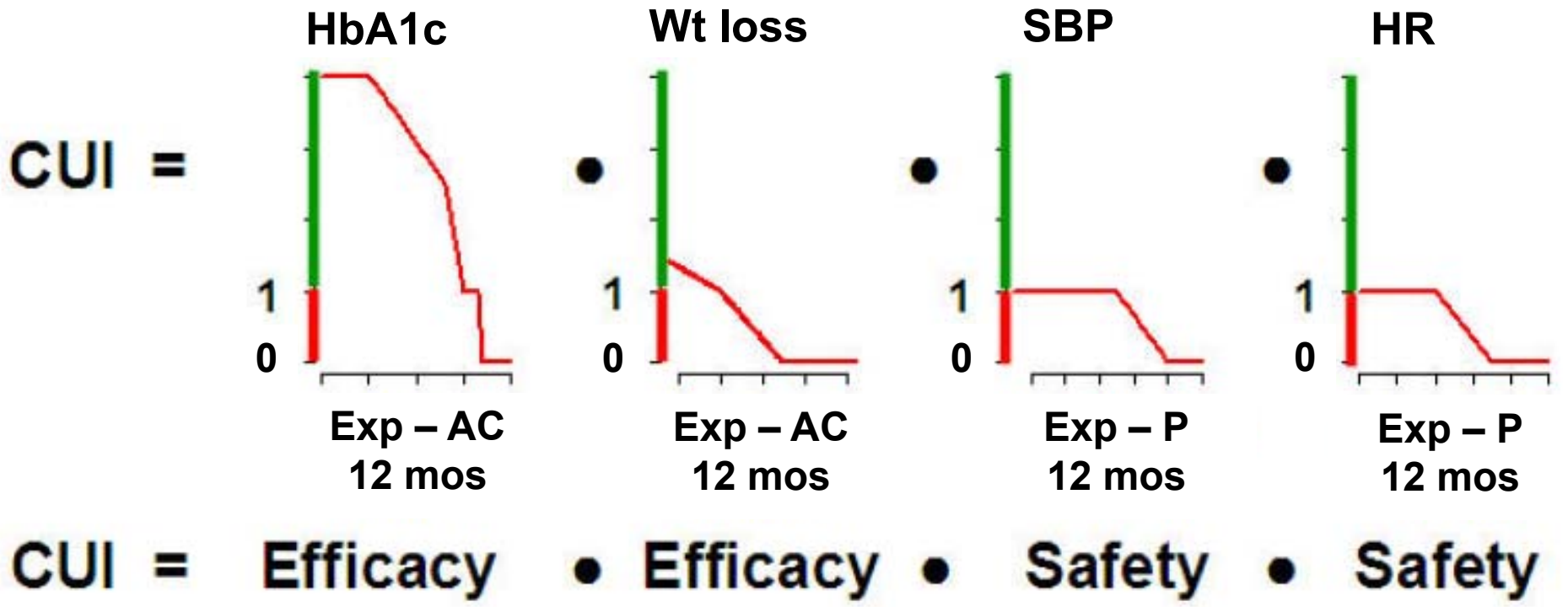
Adaptive Trial with Longitudinal Modeling

- 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

October 22, 2012

Lilly Diabetes Announces Positive Results of Phase III Trials of Dulaglutide in Type 2 Diabetes

Company Shares Top-line Results on 3 Completed AWARD Trials

INDIANAPOLIS, Oct. 22, 2012 /PRNewswire/ -- Eli Lilly and Company (NYSE: LLY) today announced positive top-line results of three completed Phase III AWARD trials for [redacted]onal, long-acting glucagon-like peptide 1 (G[redacted] as a once-weekly treatment for type 2 diabe[redacted]oints, as measured by reduction in hemoglobi[redacted]ng 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).

Chosen before any patients at 12 mos!

FDA News Release

FDA approves Trulicity to treat type 2 diabetes

For Immediate Release September 18, 2014

Release

Español

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

Inquiries

Media

Morgan Liscinsky
301-796-0397

Consumers

888-INFO-FDA

Adaptive Platform Trials …


NOVARTIS
TRONCS
AstraZeneca
Genentech
A Member of the Roche Group

Six

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

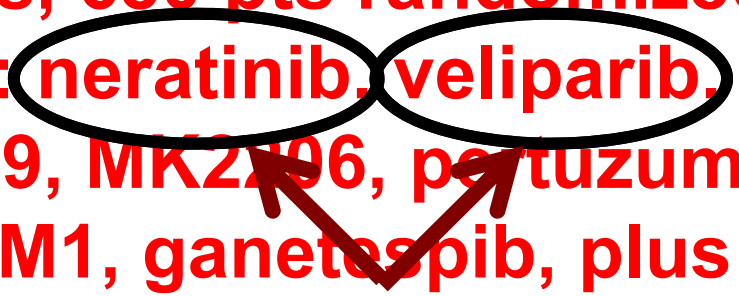
Prototype Platform Trial: I-SPY 2

http://www.ispy2.org

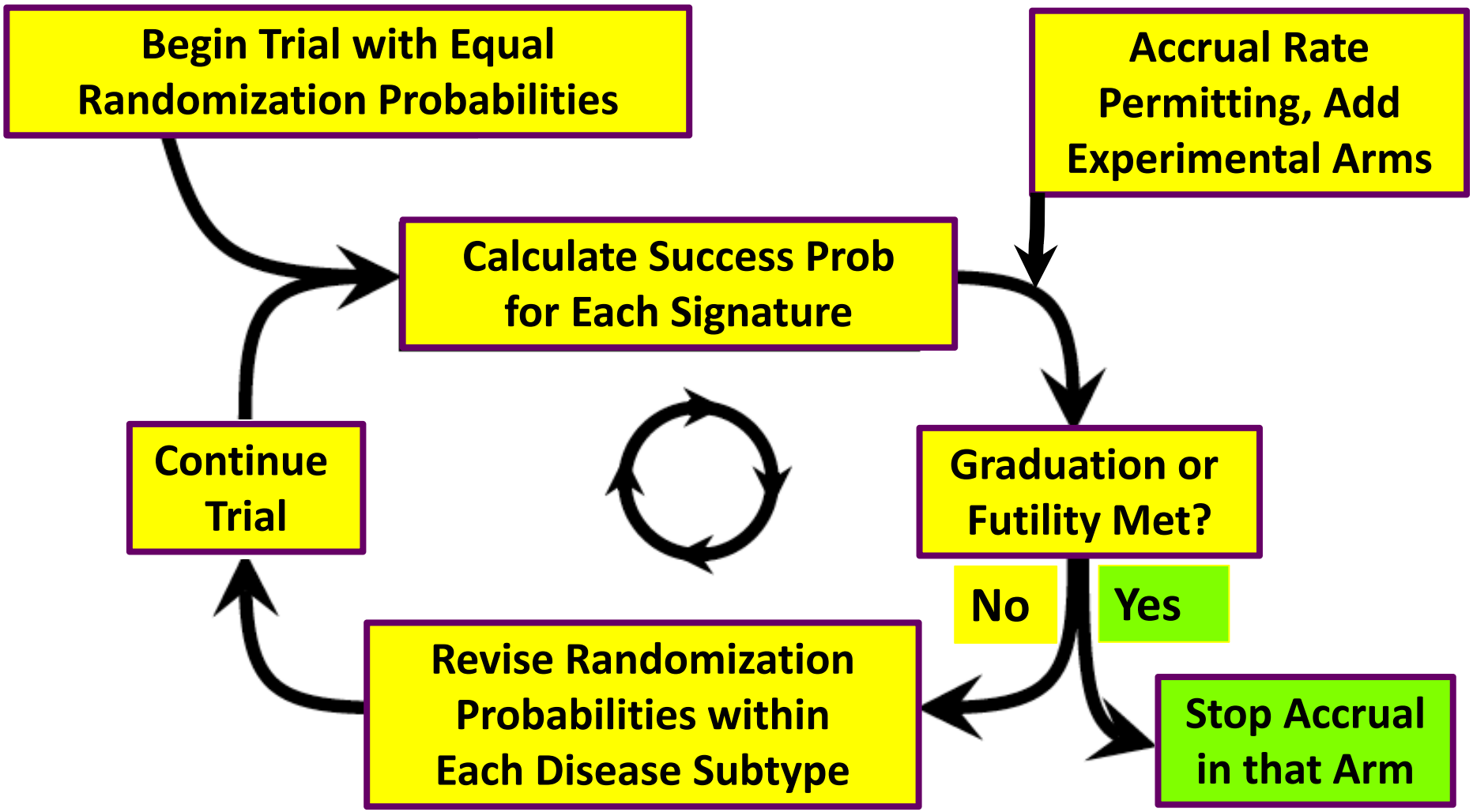
http://clinicaltrials.gov/ct2/show/NCT01042379?term=I-SPY2&rank=1

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 …



I-SPY 2 Adaptive Process

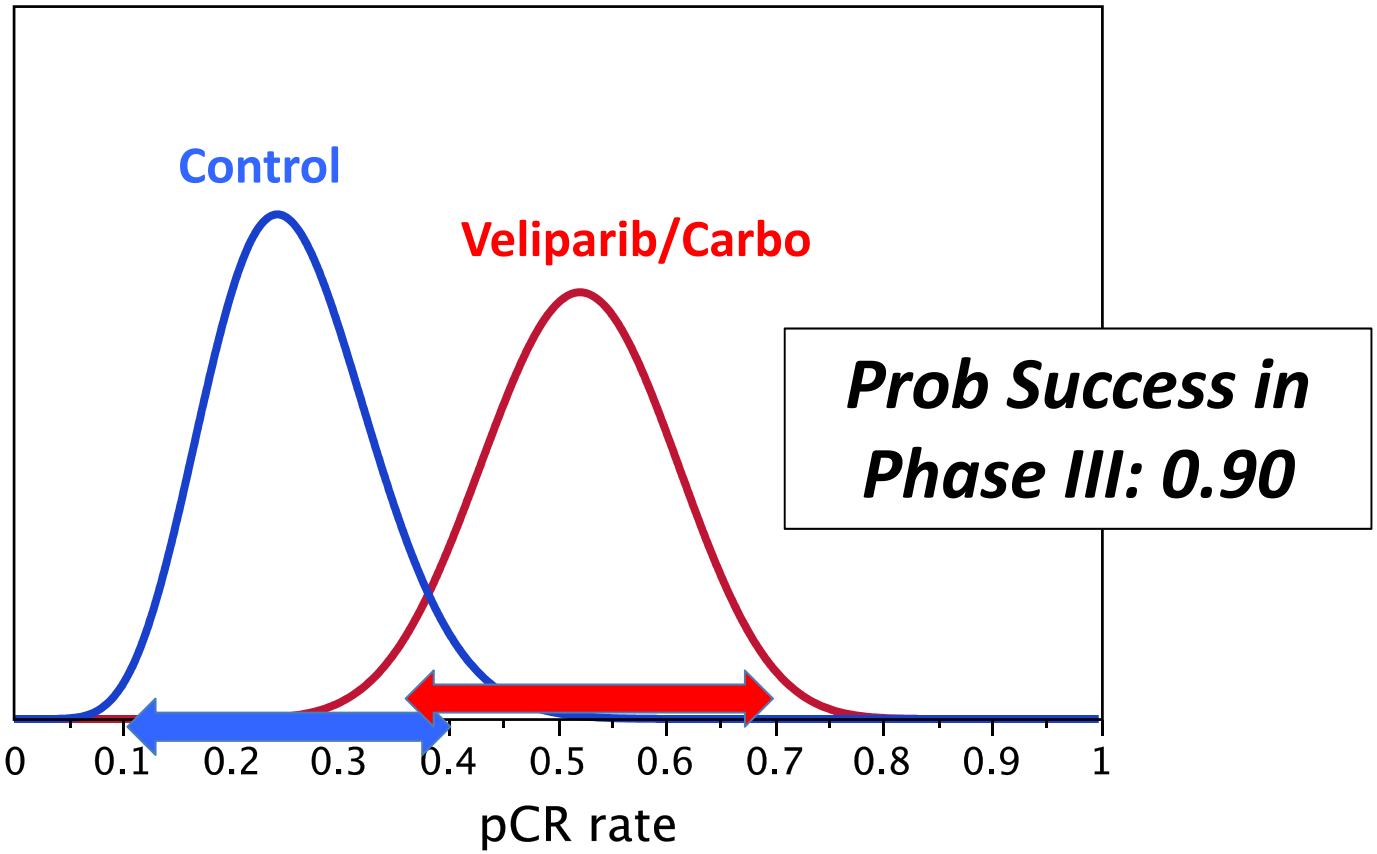


San Antonio Breast Cancer Symposium, Dec 2013

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

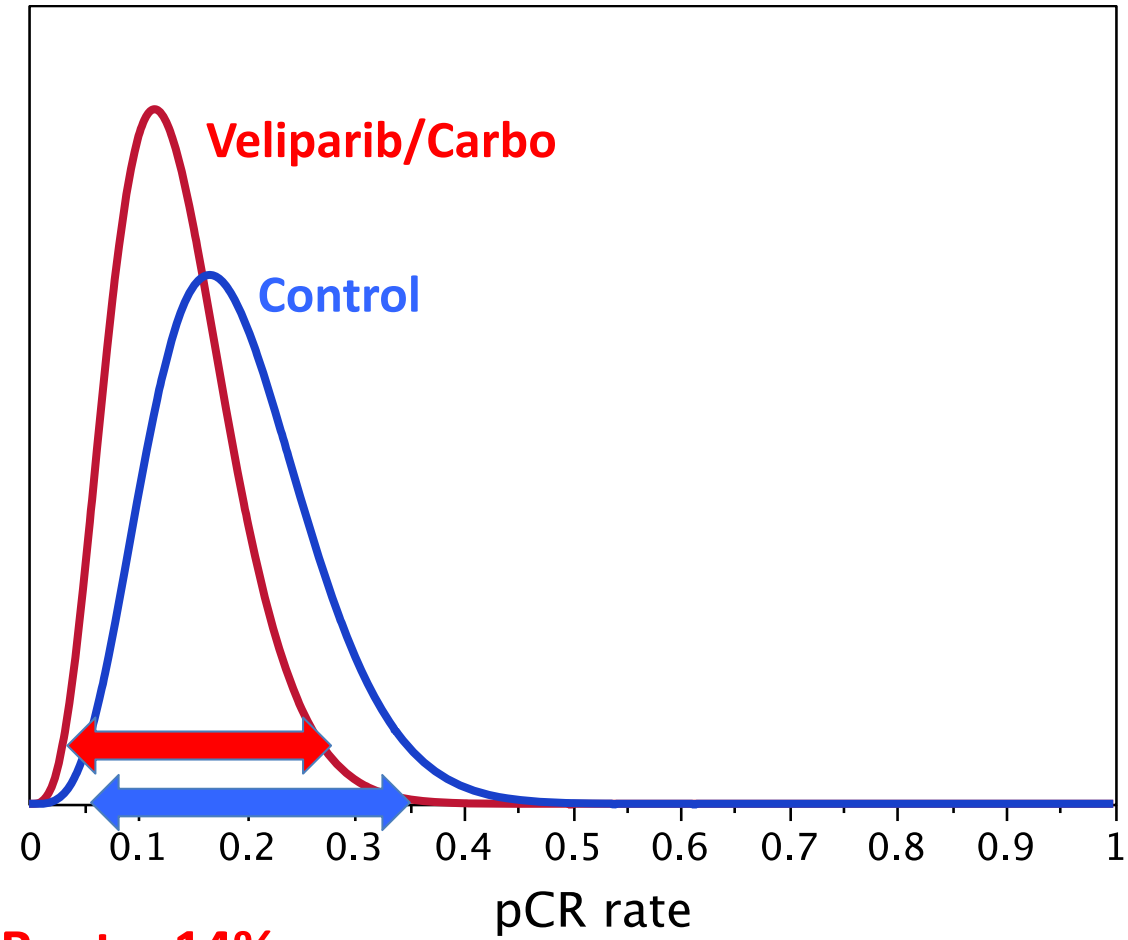
Veliparib/carboplatin graduated with Triple-negative signature



Estimated pCR rate: 26%
95% interval: 11% to 40%

Estimated pCR rate: 52%
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%

AACR 2014

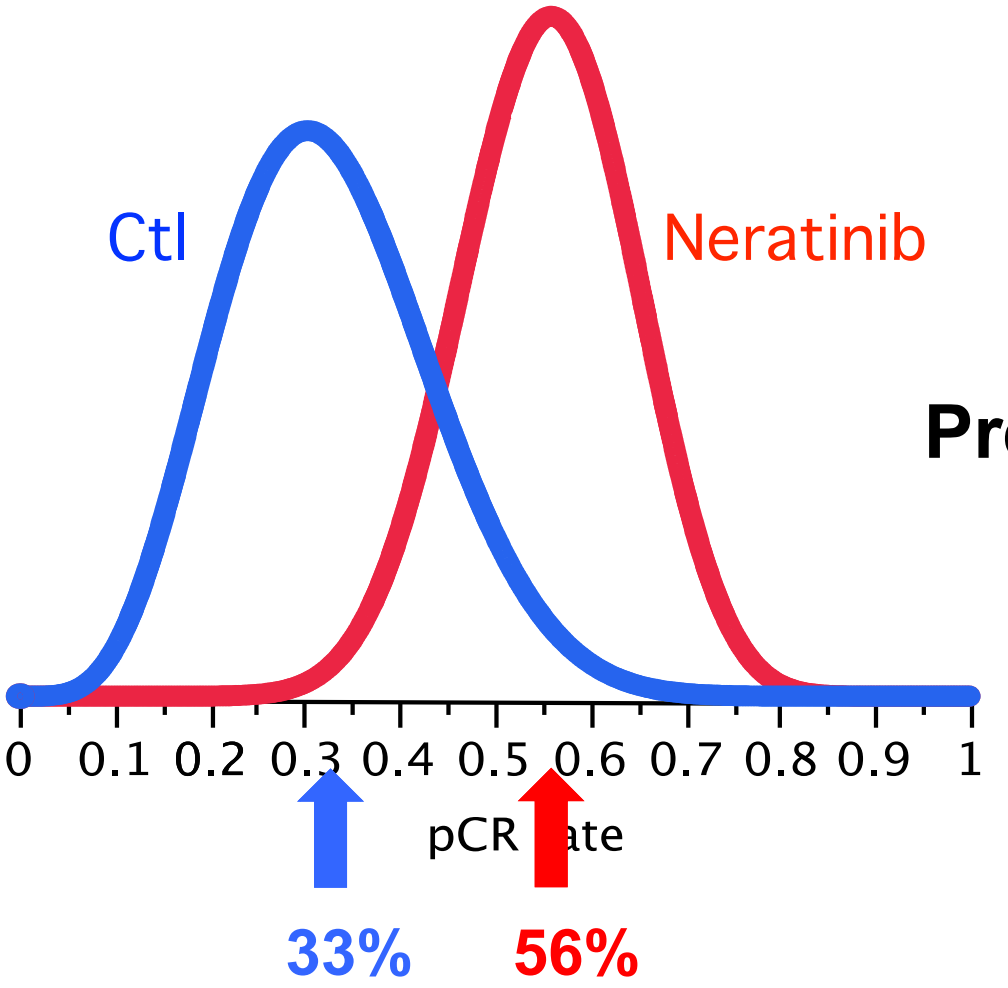
a Puma drug

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%

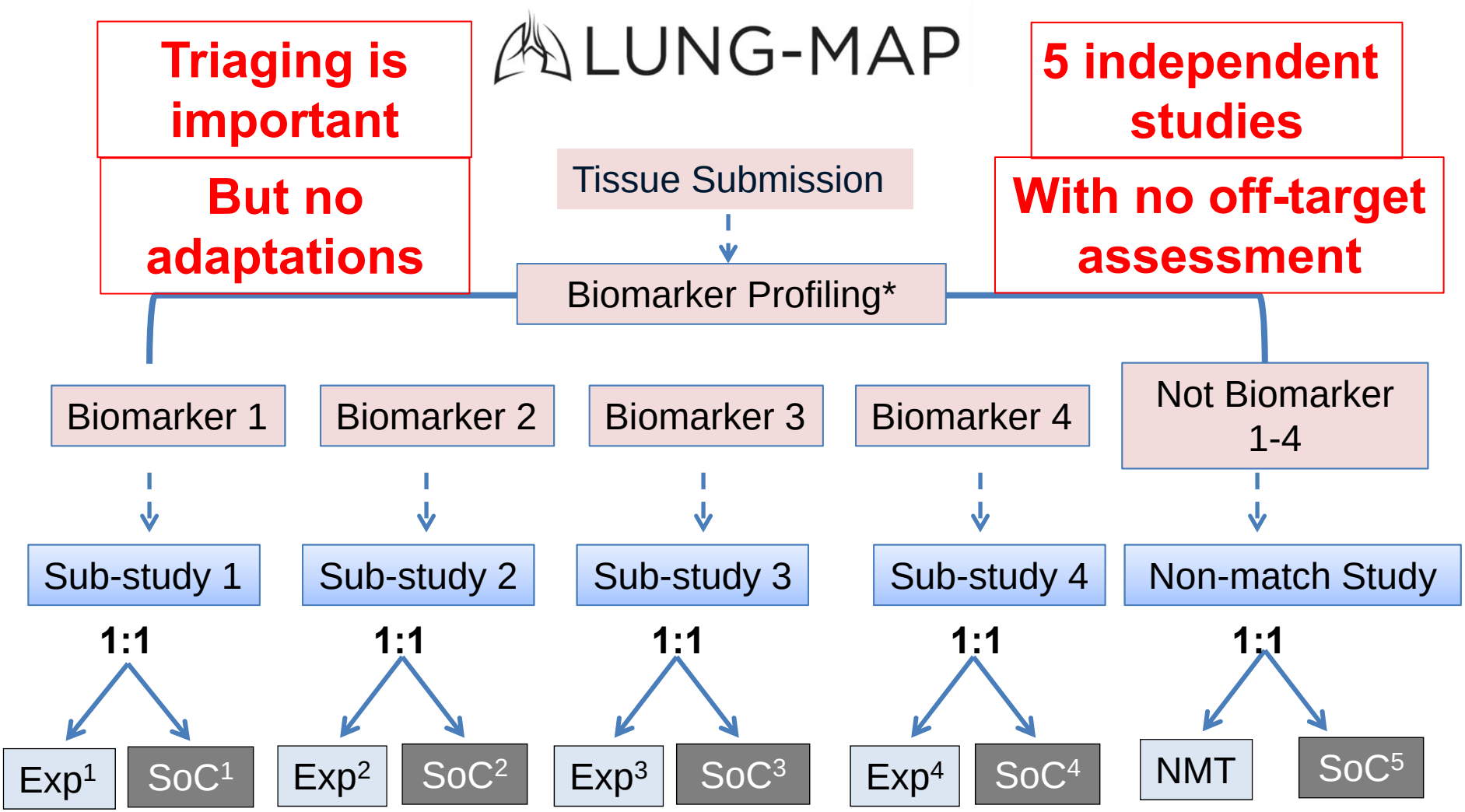
Adaptive Platform Effects

- 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+

Characteristics of Modern Platform Trials	I-SPY 2	MICAT	BATTLE	LUNG MAP	UK Matrix
Screen markers for all patients	✔	✔	✔	✔	✔
Master protocol	✔	✔	✔	✔	✔
Drugs from many companies	✔	✔	✔	✔	✔
Combination therapies	✔	✔			
Sequential therapies		✔			
Regimens enter & leave trial	✔	✔		✔	✔
Learn off-target effects	✔	✔	✔		
Pair regimens with biomarkers	✔	✔	✔		
Common control arm	✔	✔			
Adaptive randomization	✔	✔	✔		
Adaptive sample size	✔	✔			
Early “curable” disease	✔				
Registration endpoint	✔			✔	
Seamless phases				✔	
Longitudinal modeling	✔	✔			



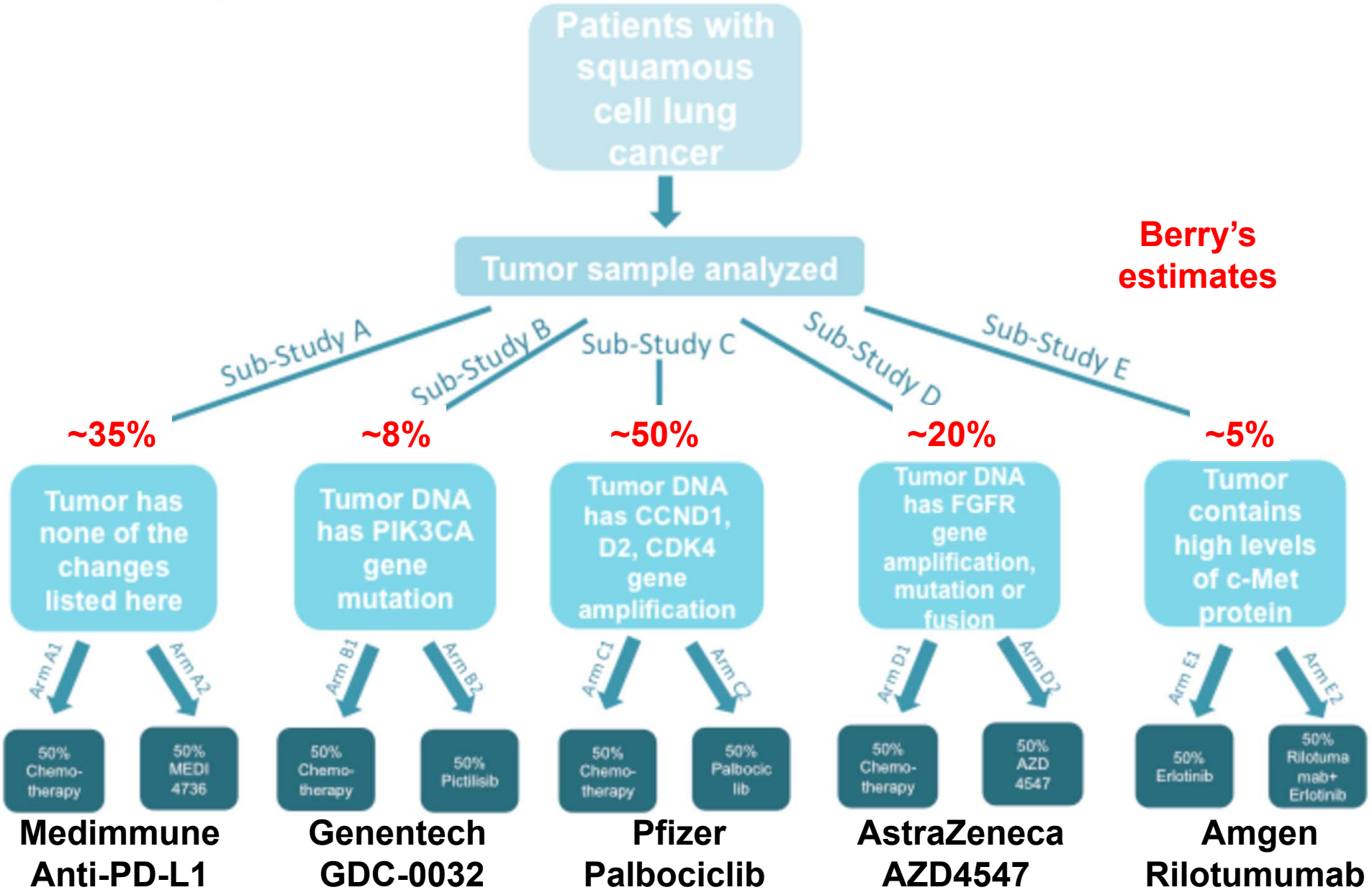
*Sub-studies assigned based on biomarker results, patients with multiple biomarkers randomly assigned to sub-study.

Exp = Targeted therapy (TT) or TT combinations (TTC), Exp^{1-4} are different TT/TTC regimens

NMT = non-match study experimental therapy or combinations

SoC = docetaxel or erlotinib, SoC^{1-5} depends on biomarker and TT/TTC/NMT regimen

Lung-MAP Sub-Studies for Treatment



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 substu
- Crossovers not allowed
- Total duration 2 – 7 yea
- 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)

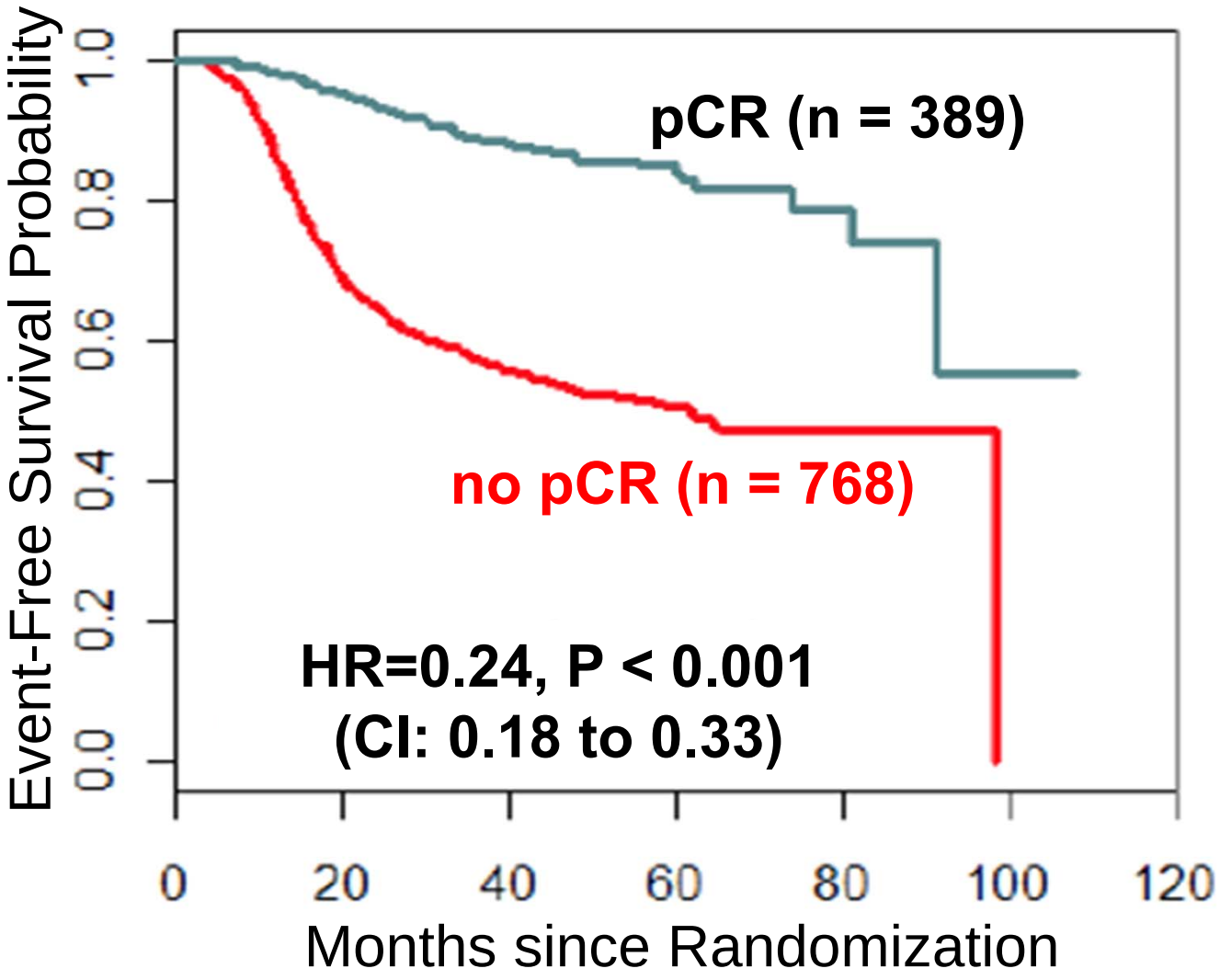
May 2012
Clinical/Medical

Efficient Phase III designs to address both pCR and EFS in a single trial

Powered for both pCR and EFS

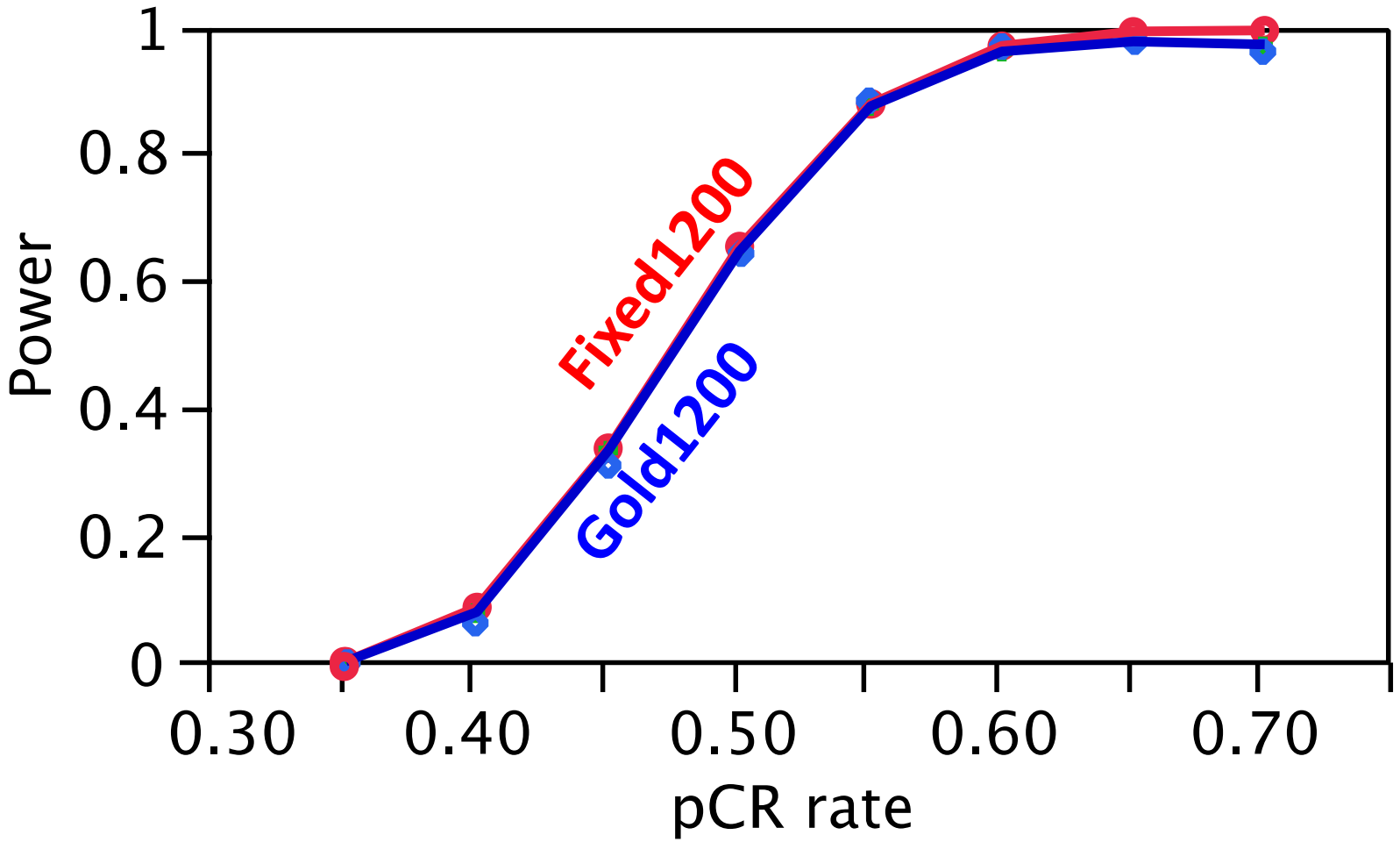
- **Accelerated approval if superiority on pCR**
  - ✓ **pCR analysis when all patients thru surgery**
- **Full approval if superiority on EFS**
  - ✓ **3 years min follow-up for EFS**
  - ✓ **Type I error rate controlled ≤ 2.5%**

Cortazar et al.
Triple-negative BC, pCR → EFS

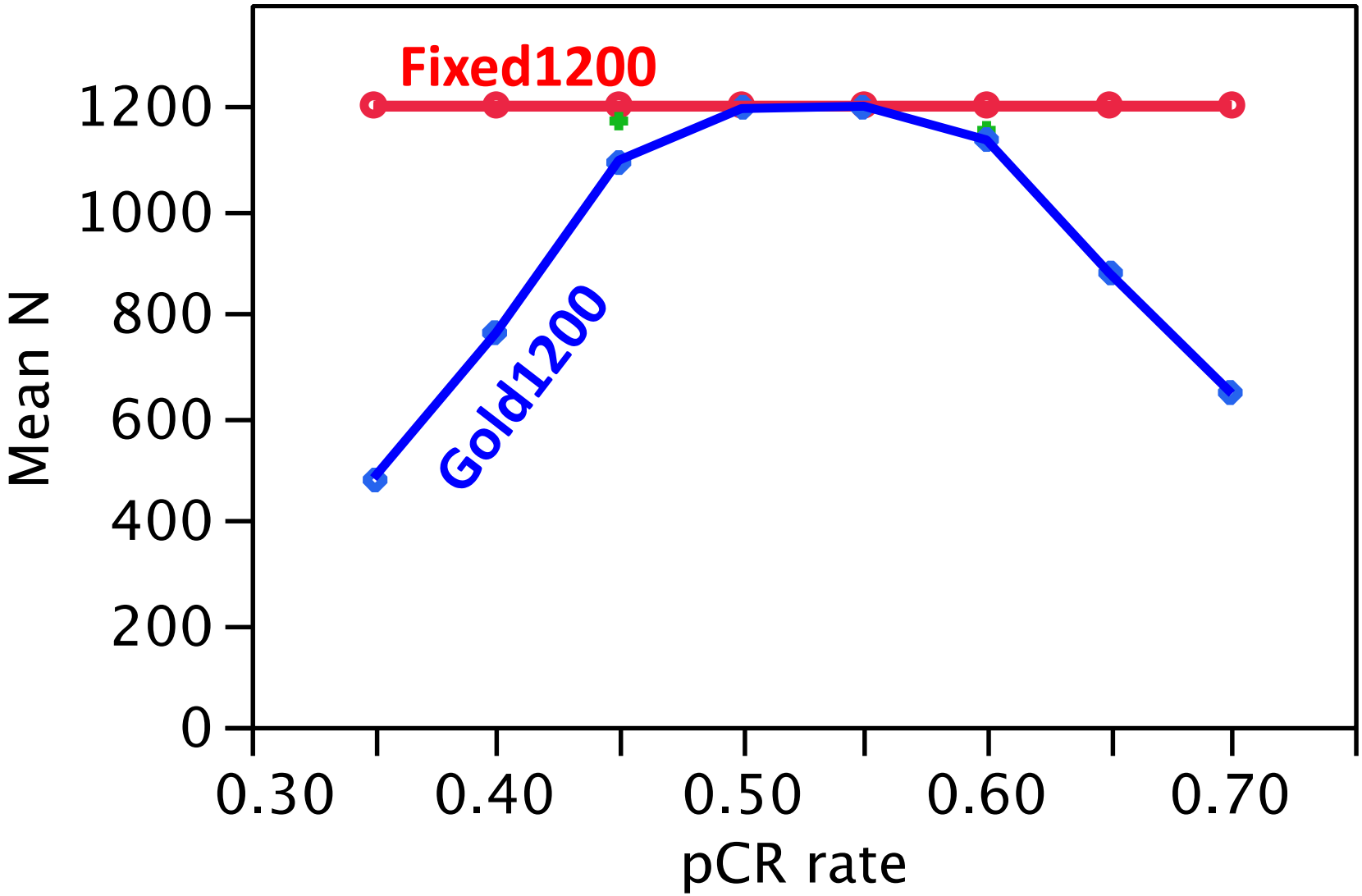


Updated information about pCR rates and relationships pCR→EFS greatly improves I-SPY 3 efficiency

Power, Goldilocks vs Fixed 1200 Design



Mean Sample Size, Goldilocks vs Fixed 1200



OUTLINE

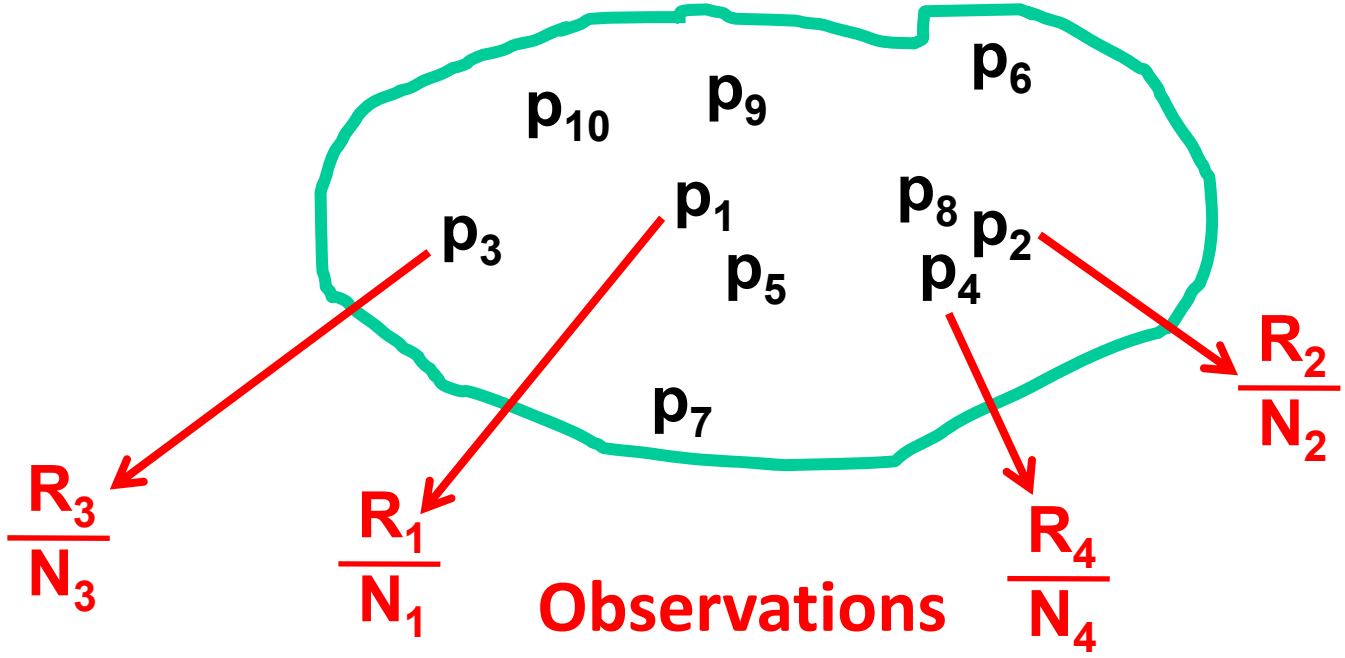
- Intro to Bayes
  - Predictive probabilities
  - Longitudinal modeling
- Adaptive trials
- I-SPY 2 (and I-SPY 3)
- **Basket trials**

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

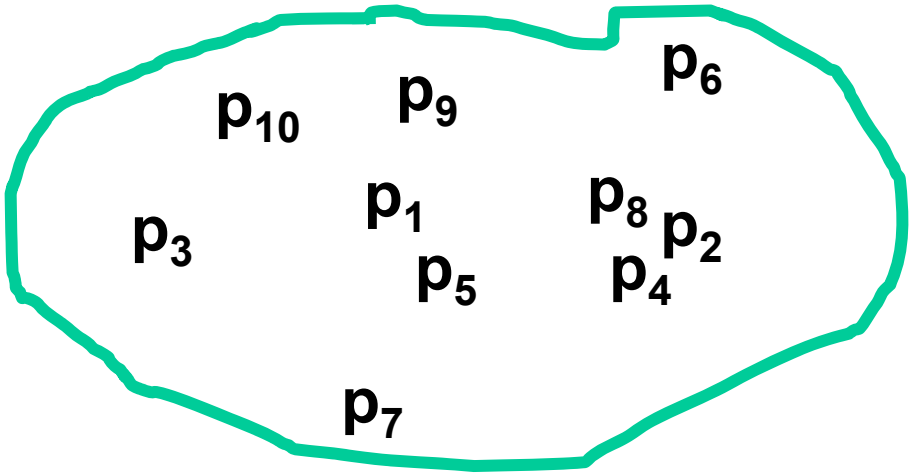
Population of response rates within tumor types:



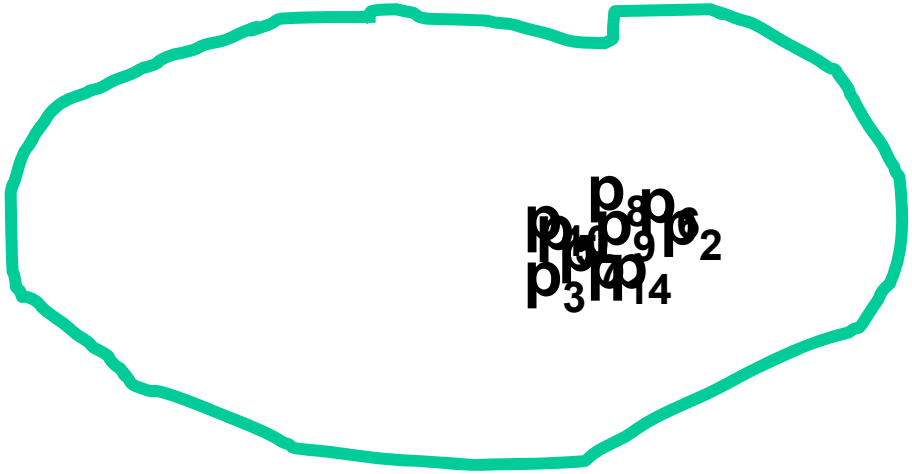
Berry SM, et al. Bayesian hierarchical modeling of patient subpopulations: Efficient designs of phase II oncology clinical trials. *Clinical Trials* 2013.

Learn about heterogeneity and clustering from trial results

Heterogeneous, little borrowing:

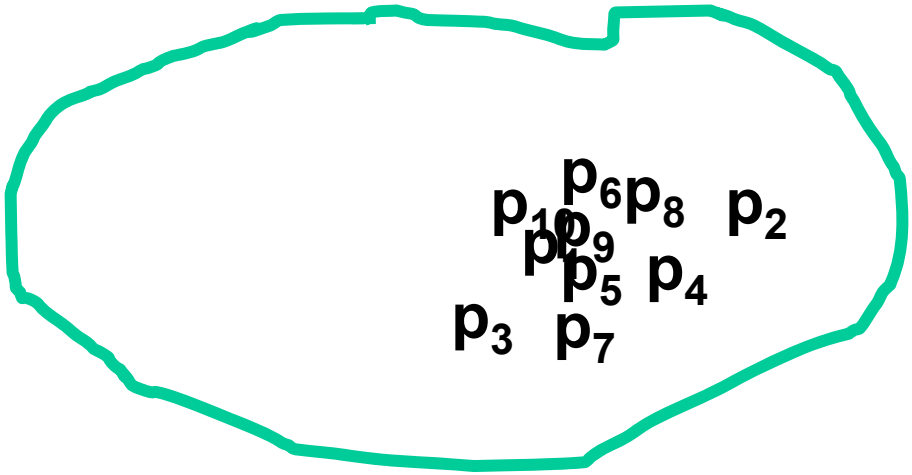


Homogeneous, much borrowing:

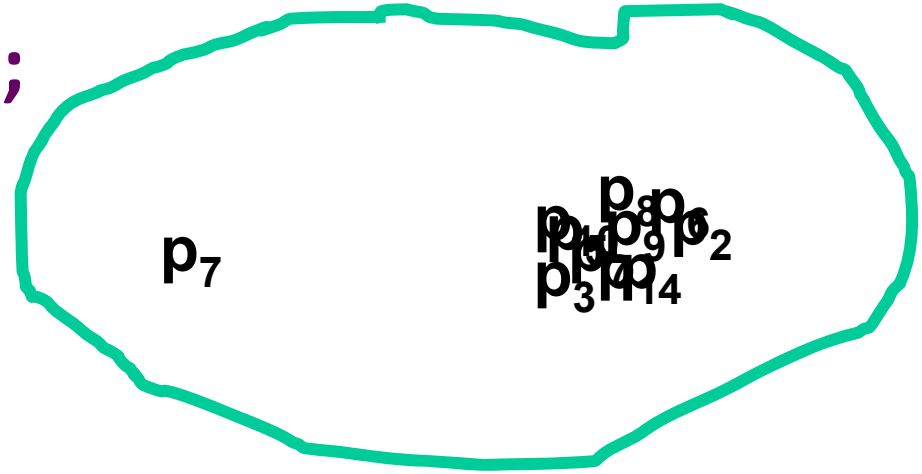


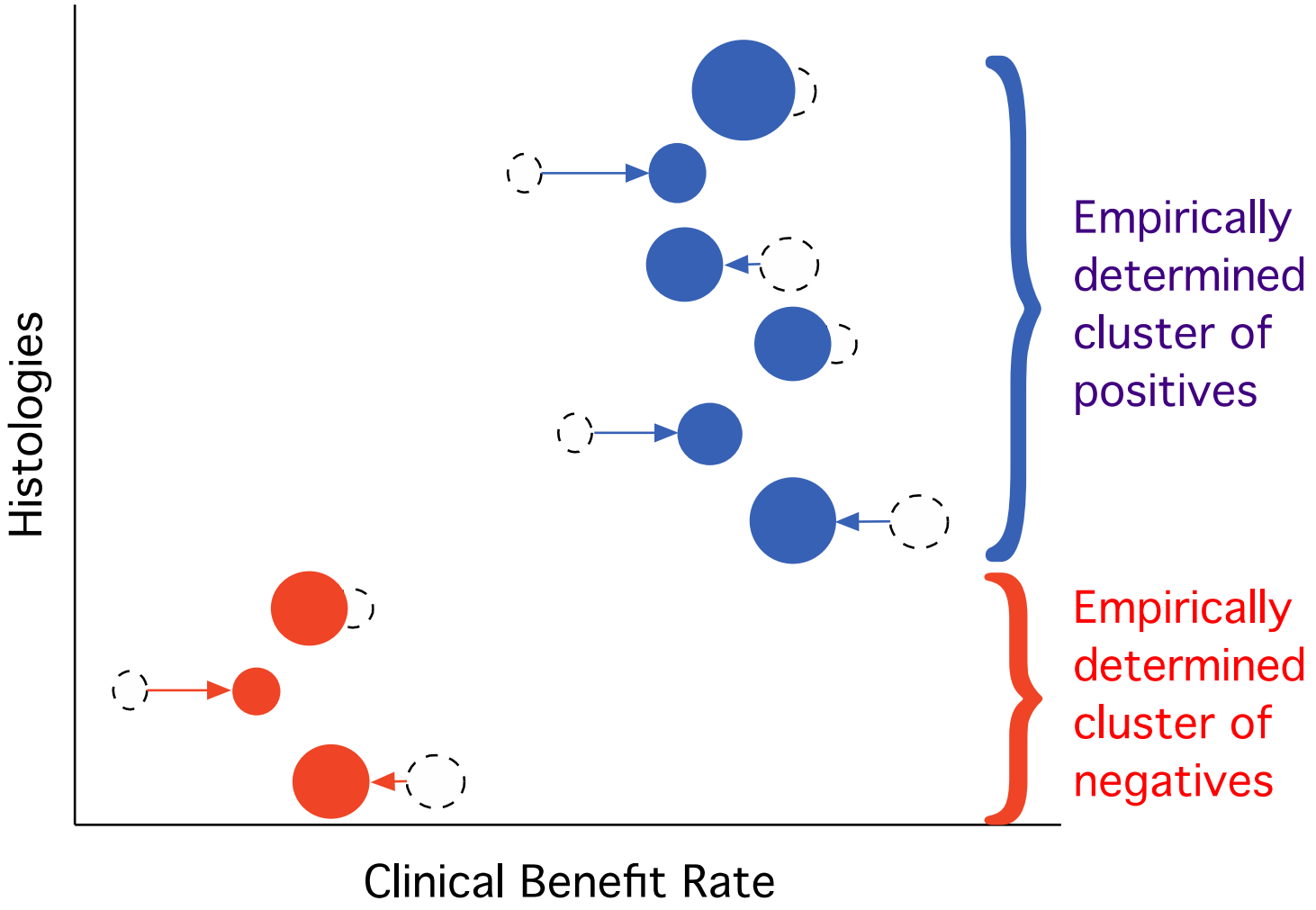
Learn about heterogeneity and clustering from trial results

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**

Signature

www.signature.com

BRING THE PROTOCOL TO THE PATIENT

Modular phase II study to link targeted therapy to patients with pathway-activated tumors[1-6]

Signature is an innovative approach to early signal finding[1-6]

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

Signature is investigating multiple pathways and compounds

		PATHWAY	COMPOUND	NCT NUMBER	
Patient preidentified with pathway-activated tumor	Study startup (3 weeks)	PI3K, PTEN[1,6]	Buparlisib (BKM120) 100 mg daily (28 days)[1,6]	NCT01833169[1]	Evaluation of primary and secondary objectives
		FGFR, PDGFR, VEGF, cKIT, FLT3, CSFR1, TrkA, RET[2,6]	Dovitinib (TKI258) 500 mg daily (5 days on, 2 days off; 28 days)[2,6]	NCT01831726[2]	
		RAF, RAS, MEK, NF1[3,6]	Binimetinib (MEK162) 45 mg bid (28 days)[3,6]	NCT01885195[3]	
		BRAFV600[4,6]	Encorafenib (LGX818) 300 mg daily (28 days)[4,6]	NCT01981187[4]	Adaptive statistical design
		PTCH1, SMO[5,6]	Sonidegib (LDE225) 800 mg daily (28 days)[5,6]	NCT02002689[5]	
		ALK, ROS1	Ceritinib (LDK378) 750 mg daily (28 days)	COMING SOON	
		CDK4/6, Cyclin D1/3, p16	LEE011 600 mg daily (3 weeks on, 1 week off)	COMING SOON	
		FGFR	BGJ398 125 mg daily (3 weeks on, 1 week off)	COMING SOON	

Treatment duration: until unacceptable toxicity, disease progression, death, and/or treatment discontinuation.[2-6]

Berry Consultants "signature"

Adaptive statistical design

Eligibility for NCT01833169

- 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