DANA-FARBER CANCER INSTITUTE

HARVARD MEDICAL SCHOOL

Bayesian hierarchical models for adaptive randomization in biomarker-driven studies: Umbrella and platform trials

William T. Barry, PhD

Nancy and Morris John Lurie Investigator
Biostatistics and Computational Biology
Dana-Farber Cancer Institute

Nov 9th, 2017

Motivation for biomarker-driven trials in oncology (in brief)

- Molecular heterogeneity of cancer is no longer a hypothesis, but known, measurable, and quantified.

  Personalized/precision medicine: A fundamental assumption is that using the genetic makeup of the tumor and the genotype of the patient will enable targeted therapeutics to improve clinical outcomes.

- Increased development of targeted therapies in oncology

- Components of multiplex genomic screening platforms are ~~converging~~ increasing overlapping

Motivation for biomarker-driven trials in oncology (in brief)

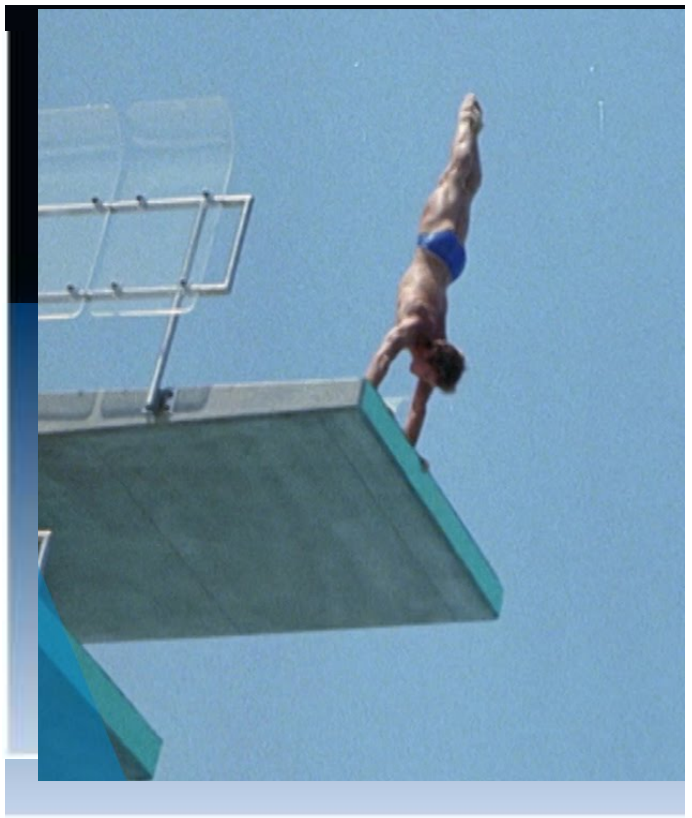
- Many innovative clinical trials designs in oncology. Important to distinguish elements:
  - Bayesian vs frequentist analysis plans
  - Comparative vs non-comparative hypotheses
  - Single-stage vs. sequential vs. continual assessment
  - Adaptive vs fixed randomization.
  - Hypotheses within or across marker-defined subgroups

Biomarker-driven designs

Integral biomarkers - Tests inherent in the design from the onset and must be performed in real time for the conduct of the trial (re: participant flow)

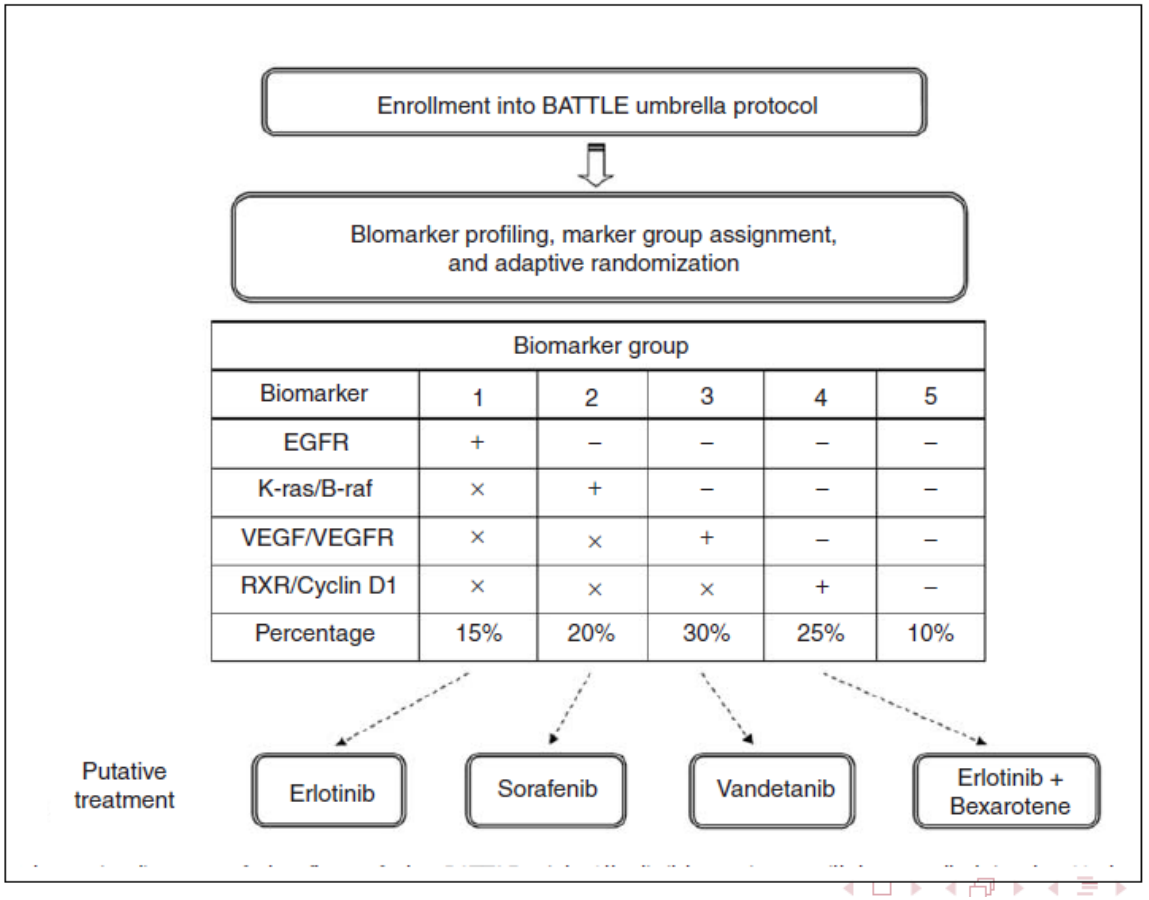
- **Single marker / treatment**
  - Enrichment designs (e.g. B31/N9831)
  - Stratified designs (TKIs and PI3Ki in Br)

- **Multiple markers / treatments**
  - Basket and Umbrella trial (BATTLE)
  - Platform trials
    - NCI-MATCH
    - I-SPY 2
  - Marker-strategy designs (SHIVA)



Herbst et al. Clin Cancer Res 2015;21:1514-1524

BATTLE: Biomarker-integrated Approaches of Targeted Therapy for Lung Cancer Elimination (PI: Kim)



Zhou et al. (2008) Clinical Trials 5:181-193 – Method (but no code) fully specified

Kim et al. (2011) Cancer Discovery 1:44-53 – Primary results

BATTLE trial design:

- Hierarchical model
- Bayesian (non-comparative) inference.
- Continual assessment
- Adaptive randomization

Kass and Steffey, 1989: Conditional independence of y_n given θ. Further, $\{\theta_{jk}|\phi\}$ are i.i.d., such that the elements of y_n are exchangeable.

$$y_n|\theta \sim p(y_n|\theta) = \prod_{jk}\prod_{i=1}^{n_{jk}} p(y_i|\theta_{jk}) \qquad \theta|\phi \sim p(\theta|\phi) = \prod_{jk} p(\theta_{jk}|\phi)$$

Binary outcome and (one possible) probit hierarchical model:

$$y_{ijk} = \begin{cases} 1 & \text{if patient } i \text{ with marker } k \text{ had a response in treatment } j \\ 0 & \text{otherwise} \end{cases}$$

$$= \begin{cases} 1 & z_{ijk} \geq 0 \\ 0 & z_{ijk} < 0 \end{cases}$$

where z_{ijk} is a latent variable that follows a Gaussian distribution.

$$z_{ijk} \sim N(\mu_{jk}, 1) \qquad \mu_{jk} \sim N\left(\phi_j, \sigma^2\right) \qquad \phi_j \sim N\left(\alpha, \tau^2\right)$$

σ^2 controls the extent of borrowing across marker groups within each treatment and α and τ^2 are the second-stage priors to the model.

BATTLE: Biomarker-integrated Approaches of Targeted Therapy for Lung Cancer Elimination (PI: Kim)

BATTLE trial design:

- Hierarchical model
- Bayesian (non-comparative) inference.
- Continual assessment
- Adaptive randomization

Futility treatment j is suspended within biomarker group k under

$$Pr\left(\Phi^{-1}(\mu_{jk}) \geq 0.5 \mid y_n\right) < 10\%$$

Efficacy likewise treatment within biomarker group (i.e. 'non-comparative')

$$Pr\left(\Phi^{-1}(\mu_{jk}) \geq 0.3 \mid y_N\right) > 80\%$$

See Zhou et al (2008) for operating characteristics w/ varying $\{\mu_{jk}\}$

BATTLE: Biomarker-integrated Approaches of Targeted Therapy for Lung Cancer Elimination (PI: Kim)

BATTLE trial design:

- Hierarchical model
- Bayesian (non-comparative) inference.
- Continual assessment
- Adaptive randomization

$$r_{jk,n} = \frac{\hat{\pi}_{jk,n}}{\sum_{w \in \Omega_{k,n}} \hat{\pi}_{wk,n}} \quad \text{where} \quad \hat{\pi}_{jk,n} = E[\Phi^{-1}(\mu_{jk})|y_n]$$

Kim (2011): *We planned to randomly assign at least the initial **80** patients equally to the 4 treatments, to allow at least 1 patient in each marker group to complete treatment, thus providing sufficient data to estimate the prior probability of [disease control]*

Barry et al. JBS 2015: The use of Bayesian hierarchical models for adaptive randomization in biomarker-driven phase II studies

Research goals:

- Evaluate properties of BATTLE (PI: Kim), as one of the first umbrella trials
- In silico simulation (R code as appendix)
- Contrast RAR and continual assessment versus traditional Simon two-stage designs

Scenarios that represent the simplest cases for using predictive biomarker(s) in a two-drug study are:

- Evaluating a novel targeted agent against a standard-of-care with a single predictive biomarker, and
- Evaluating multiple experimental agents using marker(s) selective in a complementary manner.

	Single biomarker			**Complementary biom.**	
	Marker +	Marker -		Marker +	Marker -
Trt A	θ_1	θ_0		θ_1	θ_0
Trt B	θ_0	θ_0		θ_0	θ_1

For the following illustrations: $\theta_0 = 25\%$ $\theta_1 = 50\%$

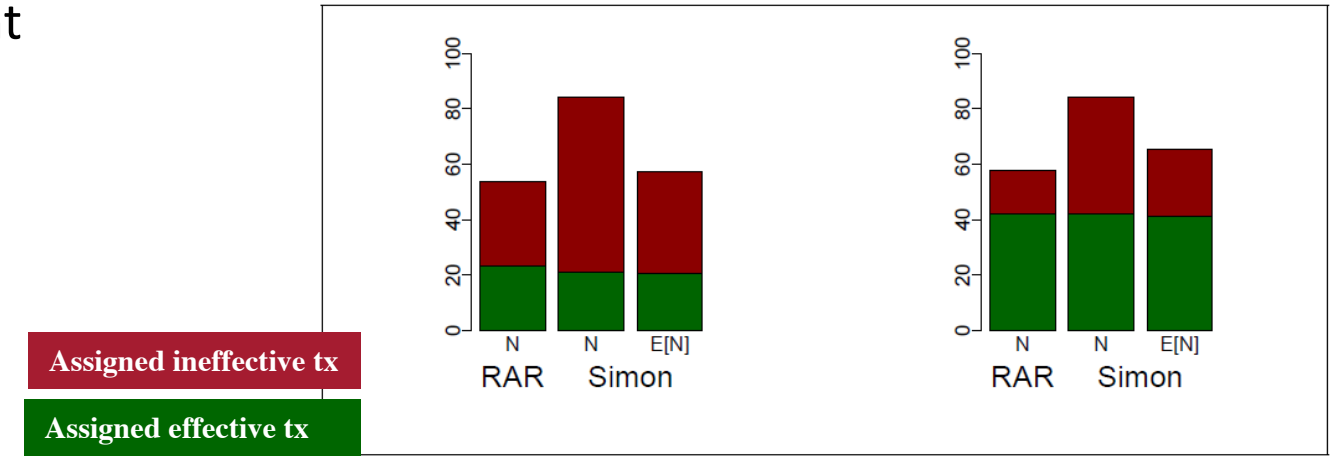
Barry et al. JBS 2015: The use of Bayesian hierarchical models for adaptive randomization in biomarker-driven phase II studies

Research goals:

- Evaluate properties of BATTLE (PI: Kim), as one of the first umbrella trials
- In silico simulation (R code as appendix)
- Contrast RAR and continual assessment versus traditional Simon two-stage designs

	Single biomarker			Complementary biom.	
	Marker +	Marker -		Marker +	Marker -
Trt A	θ_1	θ_0		θ_1	θ_0
Trt B	θ_0	θ_0		θ_0	θ_1

For the following illustrations: $\theta_0 = 25\%$ $\theta_1 = 50\%$

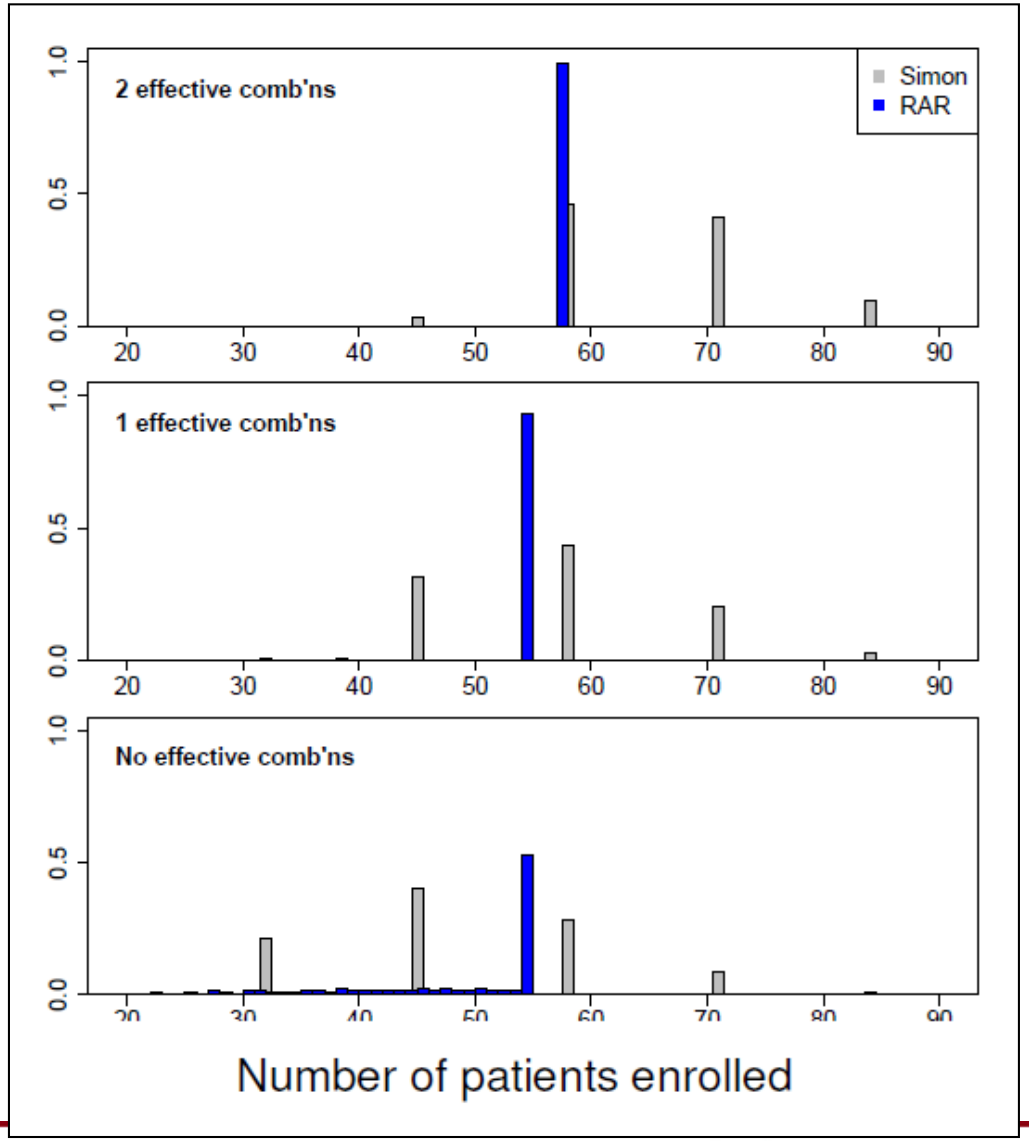


Sample sizes that achieve 80% power

Barry et al. JBS 2015: The use of Bayesian hierarchical models for adaptive randomization in biomarker-driven phase II studies

Research goals:

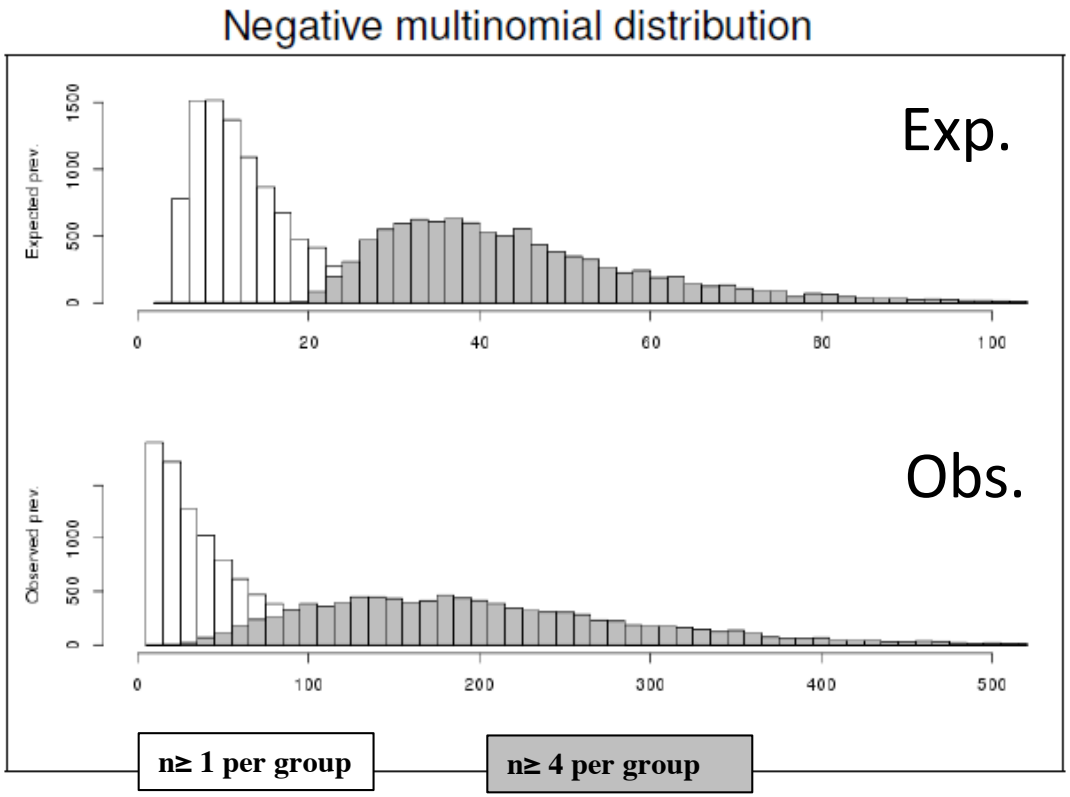
- Evaluate properties of BATTLE (PI: Kim), as one of the first umbrella trials
- In silico simulation (R code as appendix)
- Contrast RAR and continual assessment versus traditional Simon two-stage designs
- Conclusions:
  - (Nearly) equal efficiency
  - Less variability in E[N]



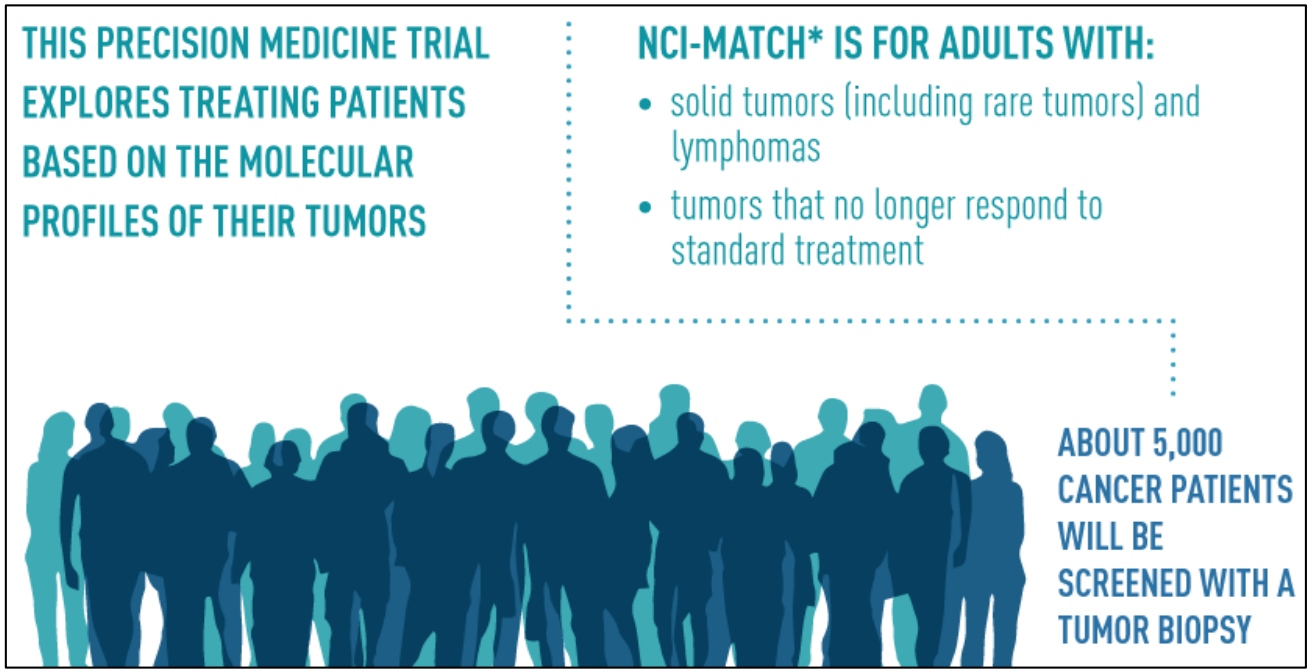
Lessons learned from BATTLE:

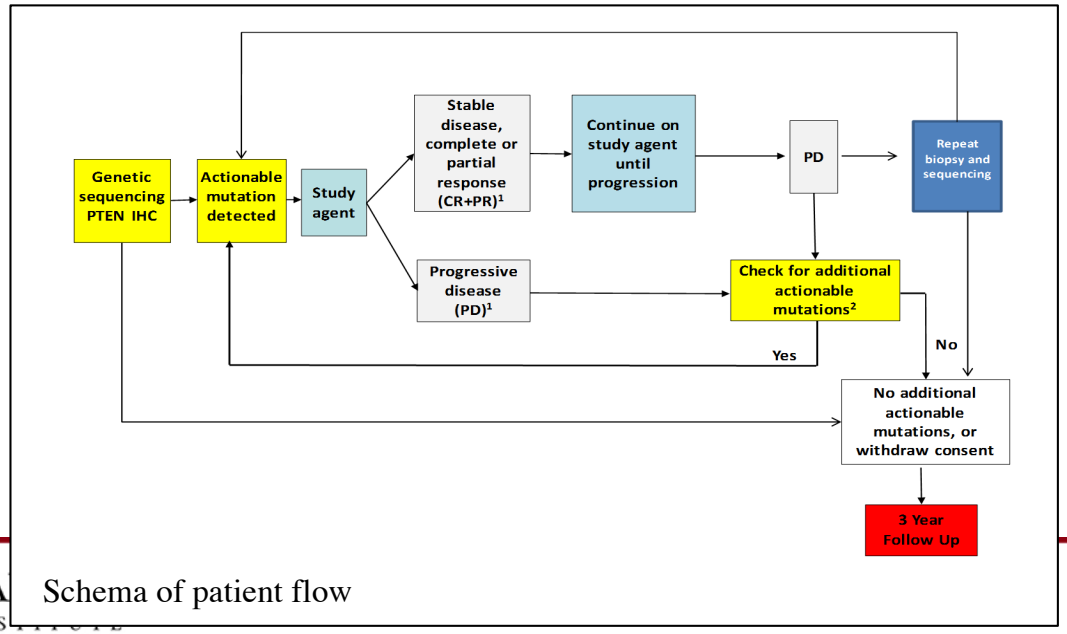
- Challenge to make reliable assumptions about prevalence of biomarkers

Group	Exp	Obs
1	10%	36%
2	20%	11%
3	30%	34%
4	25%	2%
5	10%	17%



NCI-MATCH: Molecular Analysis for Therapy Choice





Schema of patient flow

Statistical Design:

- 1° Endpoint:
  - Obj resp (RECIST1.1)
  - Null: 5%
  - Target: 25%

- Single-stage test
  - Enroll 35 pts per arm (N = 31 eval)
  - 5 or more resp.
  - $\alpha = 0.018$
  - $\beta = 0.083$

Protocol allows for expansion cohorts; not statistically driven

NCI-MATCH: Molecular Analysis for Therapy Choice

Study History	
Aug 2015	Activated with 10 initial drug arms and target N = 3000
Nov 2015	Suspended enrollment for planned evaluation 795 pts registered (739 w/ samples submitted) 645 pts completed screening 56 pts with a matching mutation (8.7%) 33 pts eligible and enrolled (5.1%) 16 pts received Tx (2.5%)
Feb 2016	Re-activated with addendum #2 Expanded eligibility to myeloma Increased to N = 5000 Increased to total of 24 treatment arms Revised estimate was 23% of pts match
Jun 2017	Reached (revised) target of N = 6000 pts 19 of 26 treatment arms still seeking patients Enrollment to sub-studies to continue through other mech's

http://ecog-acrin.org/nci-match-eay131

Snapshot of study status (Nov 2016)

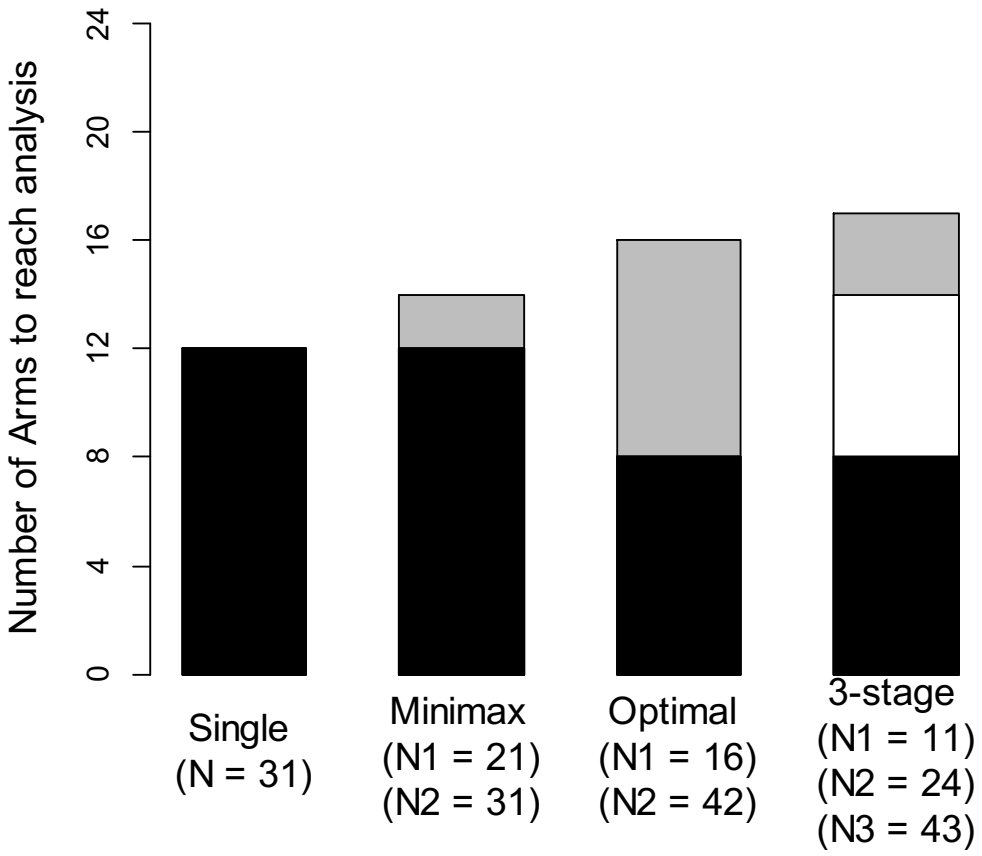
Increased
Target N: 6000 pts

24 gene alt'ns
being targeted

Arm / Target	Expected # Patients
I PIK3CA mut	137
W FGFR1/2/3	124
P PTEN loss	79
Z1A NRAS mut	70
S1 NF1 mut	66
Z1D dMMR	63
N PTEN mut	62
Q ERBB2 amp	59
B ERBB2 mut	39
C2 MET ex 14 sk	37
Z1B CCND1 amp	36
Y AKT1 mut	32

Arm / Target	Expected # Patients
R BRAF non V600	29
H BRAF V600	26
T SMO/PTCH1	18
U NF2 loss	17
C1 MET amp	14
A EGFR mut	8
G ROS1 transloc	8
S2 GNAQ/GNA11	3
E EGFR T790M	1
F ALK transloc	1
X DDR2 mut	0
V cKIT mut	0

http://ecog-acrin.org/nci-match-eay131



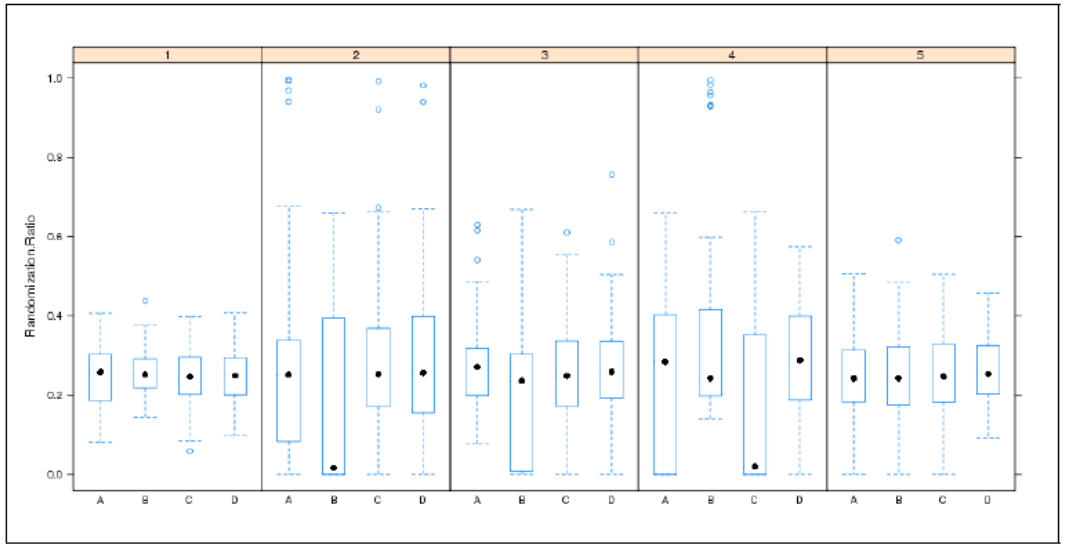
Ongoing work by R Sapigao:

- In silico simulation of the dynamic aspect of adding arms to NCI-MATCH over time and replacing completed arms
- Explore the properties of two- and three-stage designs in this framework
- Add (simulated) responses and assess Bayesian methods for continual assessment.

Lessons learned from BATTLE:

- Challenge to make reliable assumptions about prevalence of biomarkers
- Adapting w/ small n_{jk}

BATTLE: $r_{jk,n} = \frac{\hat{\pi}_{jk,n}}{\sum_{w \in \Omega_{k,n}} \hat{\pi}_{wk,n}}$ where $\hat{\pi}_{jk,n} = E[\Phi^{-1}(\mu_{jk})|y_n]$

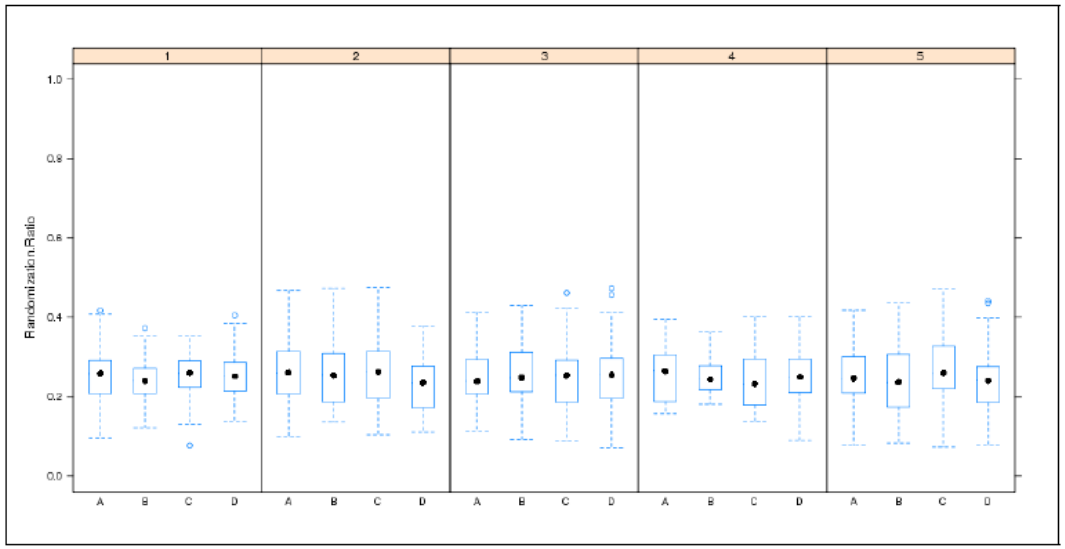


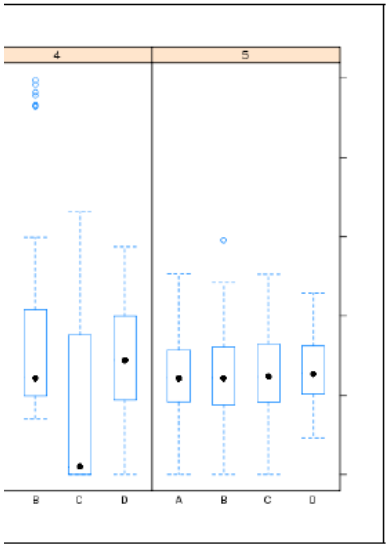
without borrowing ($\sigma^2 = 10^6$) and a non-informative prior ($\tau^2 = 10^6$).
NOTE: assuming true equipoise and no lag at $n = 97$.

Lessons learned from BATTLE:

- Challenge to make reliable assumptions about prevalence of biomarkers
- Adapting w/ small n_{jk}

BATTLE: $r_{jk,n} = \frac{\hat{\pi}_{jk,n}}{\sum_{w \in \Omega_{k,n}} \hat{\pi}_{wk,n}}$ where $\hat{\pi}_{jk,n} = E[\Phi^{-1}(\mu_{jk})|y_n]$





In Barry (in press) we advocate informative prior (e.g. $\tau^2 = 0.01$), though at $n = 97$ the likelihood will still dominate under this hierarchy.

ormative prior ($\tau^2 = 10^6$).
at $n = 97$.

Lessons learned from BATTLE:

- Challenge to make reliable assumptions about prevalence of biomarkers
- Adapting w/ small n_{jk}
- Inference w/ small n_{jk}

$$Pr\left(\Phi^{-1}(\mu_{jk}) \geq 0.3 \quad | \quad y_N\right) > 80\%$$

Table 2. Eight-week disease control status by treatment and marker groups

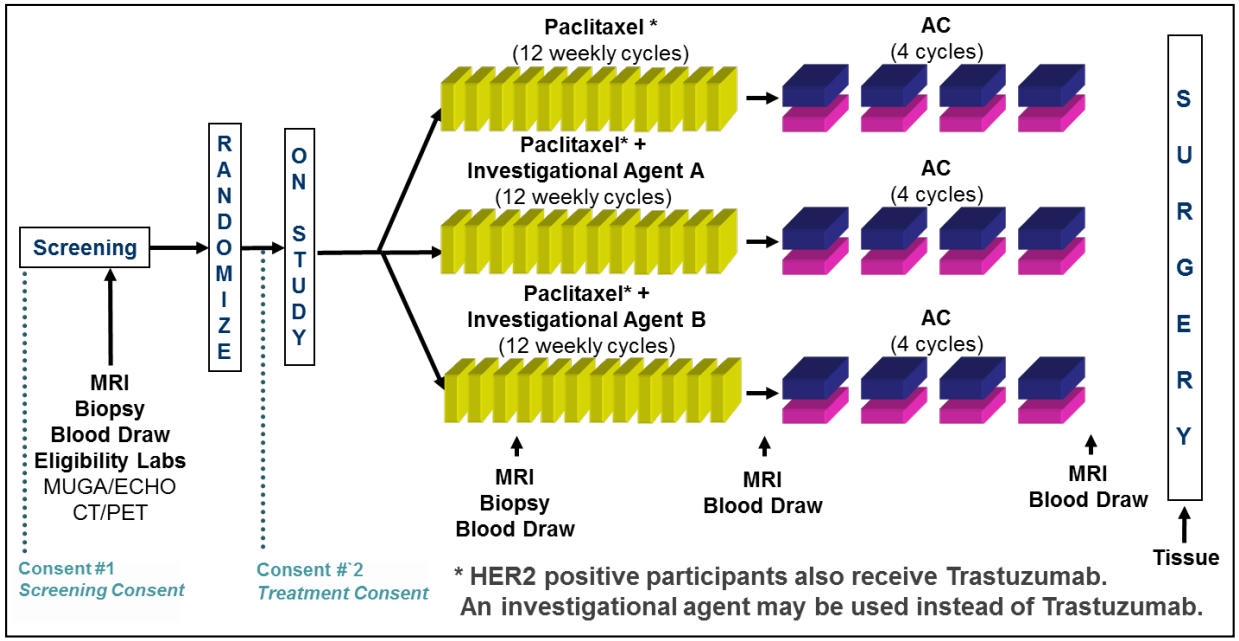
Number of patients with disease control / total number of patients (%)

Marker group	Treatment				Total
	Erlotinib	Vandetanib	Erlotinib + bexarotene	Sorafenib	
EGFR	6/17 (35%)	11/27 (41%)[a]	11/20 (55%)[a]	9/23 (39%)	37/87 (43%)
KRAS/BRAF	1/7 (14%)	0/3 (0%)	1/3 (33%)	11/14 (79%)[a]	13/27 (48%)
VEGF/VEGFR-2	10/25 (40%)[a]	6/16 (38%)	0/3 (0%)	25/39 (64%)[a]	41/83 (49%)
RXR/Cyclin D1	0/1 (0%)	0/0 (NA)	1/1 (100%)[a]	1/4 (25%)	2/6 (33%)
None	3/8 (38%)	0/6 (0%)	5/9 (56%)[a]	11/18 (61%)[a]	19/41 (46%)
Total	20/58 (34%)	17/52 (33%)	18/36 (50%)	57/98 (58%)	112/244 (46%)

[a]Cells showing effective treatments within specific marker groups defined as the probability of DCR given data is 80% or greater. Only 1 patient in the RXR/CycD1 marker group received erlotinib + bexarotene.

Likewise, $Pr > 99.9\%$ when $n_{ij} = 1$, $y_{1jk} = 1$ and $\sigma^2 \to \infty$.

I-SPY 2:
Umbrella / platform (and adaptive)



Images courtesy of Dr Rugo

Study design

- Randomized phase II
- Compare to concurrent control arm (T→AC)
- 1° endpoint: path CR
- Integral biomarkers
  - HER2
  - HR
  - Mammoprint
- Bayesian analysis plan (next slide)
- Intended to allow up to 4 experimental arms.

BATTLE trial design:

- Hierarchical model — Logistic model for pCR
- Bayesian (comparative) inference. — Threshold for 'graduation' of a regimen after 60 pts.
  Evidence (by pCR) that a future N=300 phase III study would be positive in any marker-defined subgroup: >85% PP
- Continual assessment — Threshold for futility if <10% PP in all marker-subgroups after 20 pts.

  Note: function of two parameters, π_e and π_c
- Adaptive randomization — AR is proportional to the posterior prob. a given tx is superior. Priors (appear to be) fully specified; depend on I-SPY 1

Rugo HS et al. N Engl J Med 2016;375:23-34.
Barker et al. Clin Pharmacol Ther. 2009; 86: 97– 100.

I-SPY 2: Neoadjuvant and Personalized Adaptive Novel Agents to Treat Breast Cancer

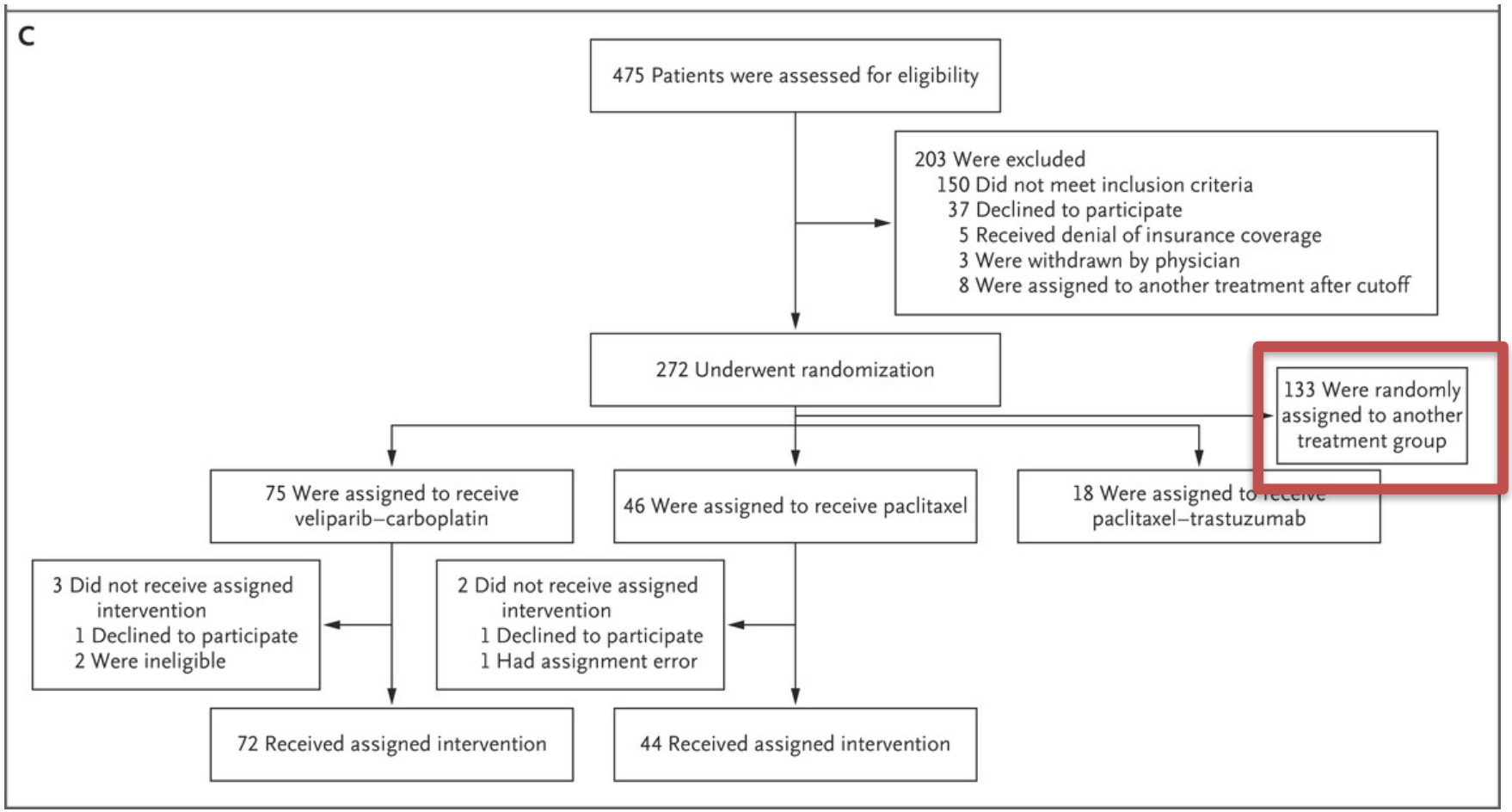
Study History	NCT00409968
Mar 2010	Activated with 3 initial experimental arms: Figitumumab, Neratinib, Veliparib + Carboplatin
Dec 2013	Results on the first regimen to 'graduate' (Veliparib + Carboplatin) were reported at SABCS by Rugo et al. **Rugo HS et al. N Engl J Med 2016;375:23-34**
Apr 2014	Results on the 2nd regimen to 'graduate' (Neratinib) were reported at AACR **Park JW et al. N Engl J Med 2016;375:11-22**
Jun 2015	Results for a 3rd regimen to 'graduate', MK-2206 [AKTi], were reported at ASCO
Jun 2017	Results for a 4th regimen to 'graduate', Pembrolizumab, were reported at ASCO

I-SPY 2: Neoadjuvant and Personalized Adaptive Novel Agents to Treat Breast Cancer

- No negative arms have been published (risk of reporting bias)
- As an ongoing study, total study-status has never been publically disseminated (to my knowledge)
- Partial information can be gleaned from clinicaltrials.gov

2010 (Target N = 800)	2012	2014	2016 (Target N = 1920)
Neratinib	Ganitumab + Metformin	AMG 386 + Trastuzumab	PLX3397
Veliparib + Carboplatin	MK-2206 +/- Trastuzumab	T-DM1 and Pertuzumab	Pembrolizumab
Figitumumab (dropped by 2012)		Pertuzumab and Trastuzumab	Talazoparib + Irinotecan
+ AMG 386		Ganetespib	Patritumab +/- Trastuzumab
+ Conatumumab (dropped by 2012)			

CONSORT: Veliparib/carboplatin



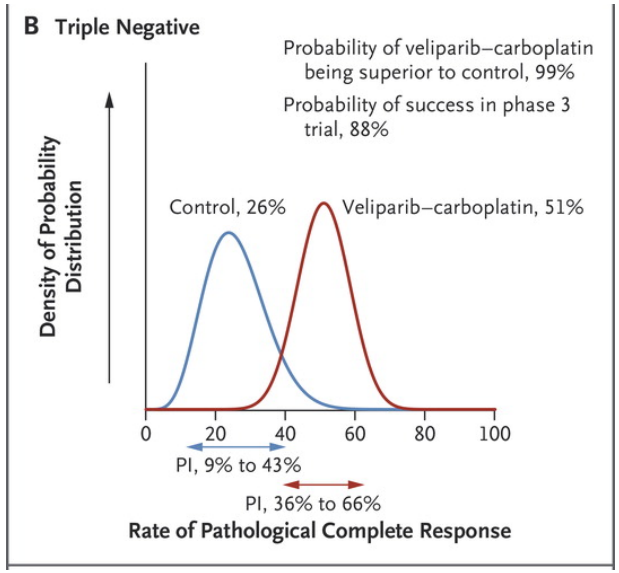
Rugo HS et al. N Engl J Med 2016;375:23-34.

Results: Veliparib/carboplatin

Table 2. Final Predictive Probabilities.*

Biomarker Signature	Estimated Rate of Pathological Complete Response (95% PI)		Probability of Veliparib–Carboplatin Being Superior to Control	Predictive Probability of Success in Phase 3 Trial
	Veliparib–Carboplatin	Control		
	percent			
All HER2 negative	33 (23–43)	22 (10–35)	91	53
Hormone-receptor positive and HER2 negative	14 (3–25)	19 (5–33)	28	8
Triple negative	51 (36–66)	26 (9–43)	99	88

* HER2 denotes human epidermal growth factor receptor 2, and PI probability interval.



"…We do not report the raw data within biomarker subtypes or signatures; our analysis carries greater precision than would a raw-data estimate"

	Veliparib + Carboplatin	Control (T → AC)
Enrolled	**N = 72**	**N = 44**
TN subset	**N = 39**	**N = 19**
pCR	20	5
No pCR	19	14

Rugo HS et al. N Engl J Med 2016;375:23-34.

* *imputed under simplified assumptions*

Comments on transparency

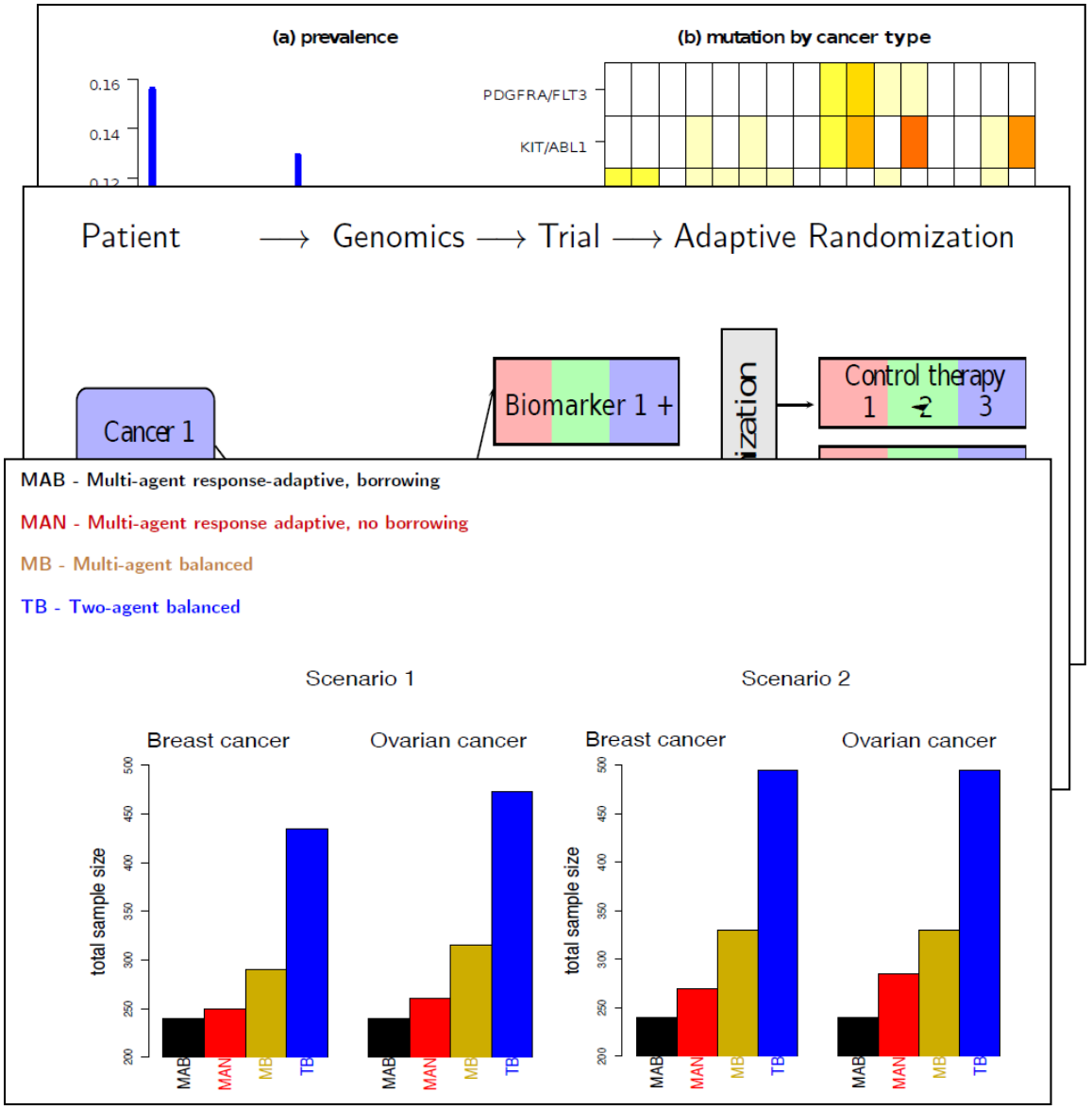
- Motivation and general approach of I-SPY 2 were published with the launch of the trial (Barker et al. 2009) Insufficient details to evaluate the specific adaptive design.
- Consistent with ICMJE policy, the protocol was provided as supplemental material to the NEJM articles.
- With multiple appendices, the statistical methods appear to be specified but will be extremely challenging to reproduce. Priors require patient-level data from I-SPY 1.
- Software has not been made public
- The decision to redact raw data from publications is concerning
- Unknown what the dissemination plans will be for negative arms

Ventz et al. Biometrics 2017: Bayesian response-adaptive designs for basket trials

Research goals:

- Develop novel methods to build off genomic platforms (Dana-Farber: Oncopanel)
- Apply RAR designs (e.g. I-SPY 2) to 'basket' trial (NCI-MATCH)
- Construct hierarchical model for adaptive allocation and continual assessment
- Use in silico simulation to tune and evaluate properties
- Provided R package(s) for models and simulation.

http://bcb.dfci.harvard.edu/~steffen/software.html



Closing remarks

- The use of master protocols (whether umbrella, basket, or platform designs) will continue to grow for trials within and across traditional disease types.
- Choice of trial design depends on many parameters:
  - Distribution of clinical outcomes, and hypothesized treatment effects
  - Marker prevalence, preliminary evidence a biomarker is predictive / prognostic, feasibility of real-time assessment, and operational resources.
- Adaptive designs give flexibility, but *always* at some cost; and it may be hard to ascertain utility
  - Response-adaptive randomization will be controversial among statisticians.
  - Adaptive enrichment designs have the potential to achieve goals of population-finding with targeted therapies.
- Adaptive platform trials are forcing us to revisit old arguments on transparency and ways to facilitate the reproducible research

Acknowledgements

Duke University and UNC
Joe Ibrahim
Chuck Perou
Lisa Carey
Kelly Marcom

Dana-Farber Cancer Institute
Steffen Ventz
Lorenzo Trippa
Giovanni Parmigiani
Rosemarie Sapigao
Meredith Regan
Richard Gelber

NCTN biostatistician:
Don Berry
Mary Redman
Bob Grey