Some efforts to embrace HTE in RCT design

Derek C. Angus, MD, MPH, FRCP

CRISMA Center, Department of Critical Care Medicine

Department of Health Policy and Management
McGowan Institute for Regenerative Medicine
Clinical and Translational Science Institute
University of Pittsburgh Schools of the Health Sciences
• The 4 dimensions of heterogeneity of treatment effect
  • Baseline probability of incurring disease-related event
  • Responsiveness to the treatment
  • Vulnerability to adverse events
  • Utilities for different outcomes

• In truth, there are several more!
Evidence-Based Medicine, Heterogeneity of Treatment Effects, and the Trouble with Averages

Richard L. Kravitz, Naihua Duan, and Joel Braslow

• The 4 dimensions of heterogeneity of treatment effect
  • Baseline probability of incurring disease-related event
  • Responsiveness to the treatment
  • Vulnerability to adverse events
  • Utilities for different outcomes

Kravitz et al. Milbank Q 2004
• The 4 dimensions of heterogeneity of treatment effect
  • Baseline probability of incurring disease-related event
  • Responsiveness to the treatment
  • Vulnerability to adverse events
  • Utilities for different outcomes
Risk of disease & risk of adverse events
Can overall results of clinical trials be applied to all patients?

P M Rothwell

• A re-analysis of the European Carotid Surgery Trial (ECST)
  • There is huge heterogeneity even within the same trial ...

<table>
<thead>
<tr>
<th>Predicted risk</th>
<th>Absolute risk reduction with surgery</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;10%</td>
<td>−1.4%</td>
</tr>
<tr>
<td>10–15%</td>
<td>7.0%</td>
</tr>
<tr>
<td>&gt;15%</td>
<td>14.1%</td>
</tr>
<tr>
<td>All patients</td>
<td>6.0%</td>
</tr>
</tbody>
</table>

Rothwell PM. Lancet 1995
What are typical distributions?

- The typical risk distribution in all clinical trials is left-shifted
- Median risk is always lower than average risk

Kent & Hayward. JAMA 2007

Knaus et al. CHEST 1991
• Assuming
  • Left-biased distribution
  • Flat hazard and constant RRR
  • Positive trial driven by large effect in minority of high-risk patients

• Implications
  • Changing net RRR across outcome risk
  • Median patient derives no benefit, and half are harmed
  • Typical subgroup analysis tests differences to the right of the median case
Drotrecogin alfa for sepsis ...

- **Multivariable subgroup analysis**
  - Despite frequency of such analyses for RCTs in NEJM, JAMA, or Lancet <1%

- **Left-bias**
  - 1,070 (63%) patients to the left of the average baseline risk
  - Benefit in a minority of high-risk patients

---

Responsiveness to treatment
Toward Smarter Lumpining and Smarter Splitting: Rethinking Strategies for Sepsis and Acute Respiratory Distress Syndrome Clinical Trial Design

Predictive Enrichment:
Selective enrollment of patients with a greater likelihood of responding to a specific therapy independent of disease severity. Greater RRR than an unselected population, and greater ARR than an unselected population (due to greater RRR).

Prognostic Enrichment:
Selective enrollment of patients with a greater likelihood of having a disease-related endpoint. Same RRR as an unselected population, but greater ARR (due to a higher event rate).

- Cystic Fibrosis G551D mutation for ivacaftor
- Lower P/F ratio for higher PEEP
- Preoperative PSA velocity for adjunctive prostate cancer treatment

Likelihood of responding to treatment

Likelihood of detecting the benefit of a treatment

Likelihood of having a disease-related event (e.g., death from ARDS)
Subphenotypes in acute respiratory distress syndrome: latent class analysis of data from two randomised controlled trials

Carolyn S Calfee, Kevin Delucchi, Polly E Parsons, B Taylor Thompson, Lorraine B Ware, Michael A Matthay, and the NHLBI ARDS Network
<table>
<thead>
<tr>
<th></th>
<th>ARMA cohort</th>
<th>ALVEOLI cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Phenotype 1 (n=318)</td>
<td>Phenotype 2 (n=155)</td>
</tr>
<tr>
<td>Mortality (at 90 days)</td>
<td>23%</td>
<td>44%</td>
</tr>
<tr>
<td>Ventilator-free days</td>
<td>17.8</td>
<td>7.7</td>
</tr>
<tr>
<td>Organ failure-free days</td>
<td>14.5</td>
<td>8.0</td>
</tr>
</tbody>
</table>

Values are estimated means that take into account the uncertainty of class membership.

*Table 4: Association between phenotype assignment and clinical outcomes, adjusted for degree of uncertainty regarding phenotype assignment*
### Table 4: Association of Phenotypes with Mortality and Changes in Outcome

<table>
<thead>
<tr>
<th></th>
<th>ARMA cohort</th>
<th></th>
<th>ALVEOLI cohort</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Phenotype 1 (n=318)</td>
<td>Phenotype 2 (n=155)</td>
<td>p value</td>
<td>Phenotype 1 (n=404)</td>
</tr>
<tr>
<td>Mortality (at 90 days)</td>
<td>23%</td>
<td>44%</td>
<td>0.006</td>
<td>19%</td>
</tr>
<tr>
<td>Ventilator-free days</td>
<td>17.8</td>
<td>7.7</td>
<td>&lt;0.001</td>
<td>18.4</td>
</tr>
<tr>
<td>Organ failure-free days</td>
<td>14.5</td>
<td>8.0</td>
<td>&lt;0.001</td>
<td>16.5</td>
</tr>
</tbody>
</table>

Values are estimated means that take into account the uncertainty of class membership.

‘Prognostic‘ – predictive of OUTCOME

### Table 4: Association of Phenotypes with Mortality and Changes in Outcome

<table>
<thead>
<tr>
<th></th>
<th>Phenotype 1 (n=404)</th>
<th></th>
<th>Phenotype 2 (n=145)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low PEEP (n=202)</td>
<td>High PEEP (n=202)</td>
<td>p value*</td>
<td>Low PEEP (n=71)</td>
</tr>
<tr>
<td>Mortality at 90 days</td>
<td>33 (16%)</td>
<td>48 (24%)</td>
<td>0.049</td>
<td>36 (51%)</td>
</tr>
<tr>
<td>Ventilator-free days</td>
<td>20 (10–25)</td>
<td>21 (3–24)</td>
<td>0.018</td>
<td>2 (0–21)</td>
</tr>
<tr>
<td>Organ failure free-days</td>
<td>22 (11–26)</td>
<td>22 (9–26)</td>
<td>0.003</td>
<td>4 (0–18)</td>
</tr>
</tbody>
</table>

Data are n (%) or median (IQR). *p value for interaction between positive end-expiratory pressure (PEEP) assignment and phenotype.

‘Predictive’ – predictive of CHANGE in outcome with therapy

Calfee et al. Lancet Respir Med 2014
Harm
2016 ‘Sepsis-3’ criteria

**SEPSIS DEFINITION**

**LIFE-THREATENING ORGAN DYSFUNCTION**

**CAUSED BY**

**DYSREGULATED HOST RESPONSE TO INFECTION**

Rapid, systemic, and complex innate response

Angus and van der Poll NEJM 2013
Disease Tolerance as a Defense Strategy
Ruslan Medzhitov et al.
*Science* 335, 936 (2012);
DOI: 10.1126/science.1214935
In silico trials of anti-TNF AB for sepsis

Pathogen load vs. Pathogen virulence
Host TNF responsiveness vs. Host anti-inflammatory responsiveness

Large difference in net effect on mortality
- from harm to benefit
- with modest swings in prevalence of unmeasured baseline variables

Clermont. Crit Care Med 2004
Sepsis ‘endotypes’

- Two patients can look similar on the outside
  - But are very different underneath

Courtesy, Tom van der Poll
So what can be done?
Strategies to ‘enrich’ based on a biomarker...

• Hope and pray models
  • Pre-specified subgroup analysis
    • Enroll everyone, only analyze a subset
  • Pre-enrich
    • Only enroll ‘biomarker positive’ patients

• Spread the bet models
  • Biomarker adaptive enrichment
  • Biomarker-stratification
Adaptive Platform Trials

• **Adaptive trials**
  • Focus on disease, not a particular Rx
  • Multiple interventions (arms)
  • ‘Perpetual’ enrollment
  • Often based on Bayes’ theorem
  • Tailor choices over time

• **So far, focused on pre-approval space ...**
  • Emphasis on efficiency with (very) small sample sizes
  • Different therapies ‘graduate’ to next phase while trial continues

Berry et al JAMA 2015
Woodcock and Lavange NEJM 2017
New Breast Cancer Results Illustrate Promise and Potential of I-SPY 2 Trial

Trial Identifies Breast Cancer Patients Likely to Benefit from Experimental Drug

By Elizabeth Fernandez on April 07, 2014

Physicians

I-SPY 2 is a collaborative research effort that uses genetic and biological markers from individual patient’s tumors to screen several promising new treatments simultaneously and allows doctors to quickly measure the effectiveness of the treatment prior to removing the tumor.

I-SPY 2 is an Innovative public-private collaboration that combines Personalized Medicine & Novel Trial Design to develop new cancer treatments much faster and for much less cost.
Adaptive Randomization of Neratinib in Early Breast Cancer


ABSTRACT

The heterogeneity of breast cancer makes identifying effective therapies challenging. The I-SPY 2 trial, a multicenter, adaptive phase 2 trial of neoadjuvant therapy for high-risk clinical stage II or III breast cancer, evaluated multiple new agents added to standard chemotherapy to assess the effects on rates of pathological complete response (i.e., absence of residual cancer in the breast or lymph nodes at the time of surgery).

METHODS
We used adaptive randomization to compare standard neoadjuvant chemotherapy plus the tyrosine kinase inhibitor neratinib with control. Eligible women were categorized according to eight biomarker subtypes on the basis of the human epidermal growth factor receptor 2 (HER2) status, hormone-receptor status, and risk according to a 70-gene profiler. All biomarkers were evaluated against control with regard to 10 biomarker signatures (prespecified combinations of subtypes). The primary end point was pathological complete response. Volume changes on serial magnetic resonance imaging were used to assess the likelihood of such a response in each patient. Adaptive assignment to experimental groups within each disease subtype was based on Bayesian probabilities of the superiority of the treatment over control. Enrollment in the experimental group was stopped when the 89% Bayesian predictive probability of success in a confirmatory phase 3 trial of neoadjuvant therapy reached a prespecified threshold for any biomarker signature (“graduation”). Enrollment was stopped for futility if the probability fell to below 10% for every biomarker signature.

RESULTS
Neratinib reached the prespecified efficacy threshold with regard to the HER2-positive, hormone-receptor-negative signature. Among patients with HER2-positive, hormone-receptor-negative cancers, the mean estimated rate of pathological complete response was 56% (95% Bayesian probability interval [PI], 57 to 79%) among 115 patients in the neratinib group, as compared with 39% among 78 controls (95% PI, 11 to 54%). The final predictive probability of success in phase 3 testing was 79%.

CONCLUSIONS
Neratinib added to standard therapy was highly likely to result in higher rates of pathological complete response than standard chemotherapy with trastuzumab among patients with HER2-positive, hormone-receptor-negative breast cancer. (Funded by QuantumLeap Healthcare Collaborative and others; I-SPY 2 Trial ClinicalTrials.gov number, NCT01043379)

---

Original Article

Adaptive Randomization of Veliparib–Carboplatin Treatment in Breast Cancer


ABSTRACT

The genetic and clinical heterogeneity of breast cancer makes the identification of effective therapies challenging. We designed I-SPY 2, a phase 2, multicenter, adaptive randomized trial to screen multiple experimental regimens in combination with standard neoadjuvant chemotherapy for breast cancer. The goal is to match experimental regimens with responding cancer subtypes. We report results for veliparib, a poly(ADP-ribose) polymerase (PARP) inhibitor, combined with carboplatin.

METHODS
In this ongoing trial, women are eligible for participation if they have stage IIC or III breast cancer with a tumor ≥2.5 cm or larger in diameter; cancers are categorized into eight biomarker subtypes on the basis of status with regard to human epidermal growth factor receptor 2 (HER2), hormone receptors, and a 70-gene assay. Patients undergo adaptive randomization within each biomarker subtype to receive regimens that have better performance than the standard therapy. Regimens are evaluated within 10 biomarker signatures (i.e., prespecified definitions of combinations of biomarker subtypes). Veliparib–carboplatin plus standard therapy was considered for HER2-negative tumors and was therefore evaluated in 3 signatures. The primary end point is pathological complete response. Tumor volume changes measured by magnetic resonance imaging during treatment are used to predict whether a patient will have a pathological complete response. Regimens move on from phase 2 if and when they have a high Bayesian predictive probability of success in a subsequent phase 3 neoadjuvant trial within the biomarker signature in which they performed well.

RESULTS
With regard to triple-negative breast cancer, veliparib–carboplatin had an 88% predicted probability of success in a phase 3 trial. A total of 72 patients were randomly assigned to receive veliparib–carboplatin, and 44 patients were concurrently assigned to receive control therapy. At the conclusion of chemotherapy, the estimated rates of pathological complete response in the triple-negative population were 53% (95% Bayesian probability interval [PI], 36 to 69%) in the veliparib–carboplatin group versus 20% (95% PI, 9 to 49%) in the control group. The toxicity of veliparib–carboplatin was greater than that of the control.

CONCLUSIONS
The process used in our trial showed that veliparib–carboplatin added to standard therapy resulted in higher rates of pathological complete response than standard therapy alone specifically in triple-negative breast cancer. (Funded by the QuantumLeap Healthcare Collaborative and others; I-SPY 2 Trial ClinicalTrials.gov number, NCT01043279)

---

**First 2 results**

- **Neratinib**
  - Tyrosine kinase inhibitor
- **Veliparib**
  - PARP inhibitor
Response-adaptive randomization

- New patient enrolled; biomarker subtype assessed
  - Randomly assign to experimental group or control group
  - For each experimental group, determine adaptive randomization probability within each subtype
  - Update probability in each experimental group vs. control for each subtype
  - Update patient outcome data
  - Update and apply longitudinal model
  - Update predictive probability for each experimental regimen vs. control in phase 3 trial for each biomarker signature

- Trial
  - Termination rule per group
    - Continue
    - Stop for futility
    - Move to next phase of trial (i.e., “graduate”)
  - Add new experimental groups if enrollment permits

Rugo et al. NEJM 2016; Park et al. NEJM 2016
A planned trial of A vs. B in 400 patients

The probability that A > B = 78%

Start randomizing MORE patients to A than B...
After 80 patients ...

Now, the probability that A > B = 99.9%
Stop the trial!
Caveats

1. If the ‘second’ 40 was flat or opposite direction ...
   • Trial continues and the next ‘bet’ swings back closer to 50:50

2. When only 2 groups, power still driven by the smaller group
   • So, NOT very helpful if ...
     • Single homogenous cohort
     • Two arms
   • But, becomes VERY interesting when ...
     • Multiple arms
     • Multiple subgroups
Response-adaptive randomization

Randomization rule

Statistical model

Px

A

B

C

DATA

Response
Response-adaptive randomization

- Odds weighted towards best RX
- Randomization rule
- Statistical model
Response-adaptive randomization

Different weights for different patient groups

Randomization rule

Statistical model
**I-SPY2: One trial with 255 potential ‘signatures’**

<table>
<thead>
<tr>
<th>MammaPrint –</th>
<th>MammaPrint+</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR+</td>
<td>HR+</td>
</tr>
<tr>
<td>HR–</td>
<td>HR–</td>
</tr>
</tbody>
</table>

- HER2+
- HER2–

- Separate treatment options and randomization probabilities within each group
- Capacity to model treatment effects across related subgroups
  - Overall and combinations of subgroups
  - I-SPY2 tests in 10 different signatures
• Trial began with uncertainty about
  • Which drugs worked
  • In whom the drugs worked

Adaptive Randomization of Neratinib in Early Breast Cancer

• Benefit in HER2+ve HR-ve patients

<table>
<thead>
<tr>
<th></th>
<th>MammaPrint –</th>
<th>MammaPrint+</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR+</td>
<td>HR–</td>
</tr>
<tr>
<td>HER2+</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td>HER2–</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Trial began with uncertainty about
- Which drugs worked
- In whom the drugs worked

Adaptive Randomization of Veliparib–Carboplatin Treatment in Breast Cancer


- Benefit in triple negative Breast Ca

<table>
<thead>
<tr>
<th></th>
<th>MammaPrint –</th>
<th>MammaPrint+</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR+</td>
<td>HR–</td>
</tr>
<tr>
<td>HER2+</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HER2–</td>
<td>0</td>
<td>++</td>
</tr>
</tbody>
</table>
Conclusions

• **Multiple axes of heterogeneity**
  • Typically consider 1 (or maybe 2) at a time ...

• ‘**Classic’ HTE: baseline risk of disease balanced against constant threat**
  • Peril: ‘one-at-a-time’ subgroups
  • Solution: large, representative trials with post-trial multivariable model

• **Precision medicine designs ...**
  • Largely IGNORE the ‘one-at-a-time’ warning
  • Instead, FOCUS on ‘predictive’ biomarkers, which may not actually predict!

• **Trial designs for putative ‘predictive’ enrichment**
  • Hope and pray
  • Spread the bet

• **Spread the bet – biomarker-enrichment**
  • Working in cancer
  • Arguably more patient-centered
CRITICAL CARE MEDICINE

CRISMA CENTER

CLINICAL RESEARCH INVESTIGATION AND SYSTEMS MODELING OF ACUTE ILLNESS

www.ccm.pitt.edu/crisma
Does tyrosine kinase inhibition help?
Does PARP inhibition help?