Although molecular profiling is expensive, not doing so is ultimately far more expensive and gives the wrong answer

(Stewart et al., JCO 2010)
Key Questions

- How do we use markers to select patients for therapy?
- How do we validate that using the marker actually helps?

Biomarker Driven Trials: Issues to Consider

- Strength of pre-clinical evidence of the marker
  - Restrict patients based on marker status or enroll all patients regardless of the marker status?
- Prevalence of the marker
  - Low / moderate / high
- Reproducibility and validity of assays
  - Local versus Central Testing
- Ethical issues in obtaining multiple biopsies from patients to assess marker status
- Feasibility and timing of assessments
## Initial Validation: Phase II Testing

### Design Considerations

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Design</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Enrichment</td>
</tr>
<tr>
<td><strong>Preliminary Evidence</strong></td>
<td>Yes</td>
</tr>
<tr>
<td>1. Strongly suggest benefit in marker defined subgroups.</td>
<td></td>
</tr>
<tr>
<td>2. Uncertain about benefit in overall population versus marker defined subgroups</td>
<td>--</td>
</tr>
<tr>
<td><strong>Assay Reproducibility and Validity</strong></td>
<td></td>
</tr>
<tr>
<td>1. Excellent (high concordance between local and central testing, commercially available kits, etc.)</td>
<td>Yes</td>
</tr>
<tr>
<td>2. Questionable</td>
<td></td>
</tr>
<tr>
<td><strong>Turnaround Times</strong></td>
<td></td>
</tr>
<tr>
<td>1. Rapid (2-3 days, without causing delay in the start of therapy)</td>
<td>Yes</td>
</tr>
<tr>
<td>2. Slow to Modest (one week or more)</td>
<td></td>
</tr>
</tbody>
</table>

Mandrekar & Sargent, JTO 2011
Prevalence of the Marker

- Low (<10 - 20%): Enrichment
- High (>50%): All-comers with retrospective marker subgroup assessments or adaptive designs
- Moderate (20%-50%): 
  - Stratified by marker, primary hypothesis in marker (+)
  - Let’s discuss more…

Possible design strategy for moderate prevalence marker

- Phase I: No restrictions
- Phase IIa (optional): Single arm, enriched
  - Proof of concept
- Phase IIb: Randomized phase II unselected
  - Primary comparison: Marker (+)
  - Randomize enough Marker (-) to demonstrate lack of benefit
  - Consider adaptive design
- Phase III: Based on randomized phase II
Adaptive Phase II Designs

- Randomize between at least 2 arms within biomarker-defined strata
  - Different signatures, different allowed drugs
- Evaluate success in an ongoing manner
  - Alter randomization ratio?
- Drop poor performers
- ‘Graduate’ good performers to phase III trials
- ISPY-2 (Breast - ongoing), BATTLE-1 (NSCLC - completed)

Zhou et al., 2008; Barker et al., 2009

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

Design Features:
- Patients classified into five biomarker subgroups
- Equal randomization in the first stage for model development
- Adaptive randomization applied after enrolling at least one patient in each marker by treatment subgroup.
- Biopsy mandated trial; pre-specified hypotheses regarding biomarkers and targeted treatments
BATTLE TRIAL

Four Molecular Pathways and Four Putative Targeted Therapies

Biomarker Profiles: $2^4 = 16$ marker groups

16 marker groups x 4 treatments = 64 combinations

Endpoint: Disease Control Rate (DCR) at 8 weeks

**BATTLE TRIAL**

Enrollment into BATTLE Umbrella Protocol

Biomarker Profile and Adaptive Randomization

<table>
<thead>
<tr>
<th>MG</th>
<th>EGFR</th>
<th>K-ras and/or B-raf</th>
<th>VEGF and/or VEGFR</th>
<th>RXR and/or cyclin D1</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>+</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>0.15</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>+</td>
<td>x</td>
<td>x</td>
<td>0.2</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>x</td>
<td>0.3</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>0.25</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.1</td>
</tr>
</tbody>
</table>

**Endpoint: Disease Control Rate (DCR) at 8 weeks**
BATTLE Trial Results

- Trial accrued a total of 255 patients-
  - 97 patients randomized equally between the 4 arms
  - 158 were adaptively randomized

- Overall, the DCR was 46% at 8 weeks; median OS = 9 months; median PFS = 1.9 months
  - DCR was 71% for patients with EGFR mutations treated with Erlotinib
  - DCR was 61% for patients with KRAS mutations treated with Sorafenib

Designs for Predictive Marker Validation
Predictive Marker Validation

- Goal: Determine which treatment will work for which patient

- Vital: Patients treated with treatment choices in question must be comparable
  - Changes in patient populations based on biologic sub-setting
  - Evolution in imaging technologies

- Only true assurance: Patients randomized between treatments in question – need for a randomized controlled trial (RCT)

Lack of Randomization!!!

- Observational study of 656 consecutive patients

- Tested association of biomarker with chemotherapy benefit

- Appears pts with marker get big therapy benefit (compare dotted lines)

- Problem: Non-randomized: Treated pts median 13 years younger than untreated!
Prospective Designs for Marker Validation

- Enrichment Designs
- Hybrid Designs
- All-comers or Unselected Designs
- Adaptive Designs

Mandrekar and Sargent, JCO 2009

Enrichment Designs

- Screens patients for the presence or absence of a marker or a panel of markers, AND

- Only includes patients in the clinical trial who either have or do not have a certain marker characteristic or profile

- Paradigm: Not all patients will benefit from the study treatment under consideration

- Goal: Understand clinical benefit in subgroup of the patient population defined by a specific marker status
Enrichment Designs

Appropriate when:

- Mechanism of drug action is known
- Assay is reliable
- Compelling preliminary evidence
- Rapid turn around times

Sample size:

- Needs fewer overall randomized patients compared to an “unselected” design
- Depends on prevalence / assay accuracy

Using markers to restrict trial eligibility: success – Her 2+ Breast Cancer

Disease-Free Survival: Joint Analysis

Romond et al, NEJM 2005
Using markers to restrict trial eligibility: beware

- Herceptin in Her2- breast cancer?
- Discordance between local and central testing for HER2 status
  - Herceptin therapy may benefit a potentially larger group than the approximately 20% of patients defined as HER2 positive by central testing in these two trials

Paik et al, NEJM 2008

- No difference in benefit based on strength of HER2+
- May need new study of Herceptin in Her2- patients!!
Hybrid Designs

- Marker defined subgroup of patients randomized between treatments,
  - Patients in the other marker subgroups assigned standard of care.
- Appropriate when
  - Compelling prior evidence demonstrating the efficacy of a treatment for a marker defined subgroup
  - Unethical to randomize patients with that particular marker status to other treatment options

Hybrid vs. Enrichment Designs

**Hybrid**
- Powered to detect differences in outcomes only in the marker defined subgroup that is randomized to treatment choices based on the marker status
- Include and collect specimens and follow-up from “all” patients in the trial

**Enrichment**
- Same as Hybrid
- Excludes patients who do not have the specific marker status
TAILORx Trial Design

Pre-Registration
(Tumor block submission)

Oncotype Dx risk score (RS)

Intermediate risk
(RS: 11-25)

Low risk (RS < 11)

High risk (RS > 25)

Randomization (1:1)

Hormonal therapy

Chemotherapy + Hormonal therapy

Hormonal therapy

Chemotherapy + Hormonal therapy

Unselected / All-comers Designs

- **Marker by treatment interaction design**
  - Use the marker status as a stratification factor when randomizing subjects to treatment

- **Subset Analysis, if overall effect is not significant**

  All patients of a specific disease type and stage are eligible for the clinical trial, regardless of their actual marker status

RS = recurrence score
Unselected Design: Upfront Stratification by Marker status

Register → Test Marker

Marker Level (-) → Randomize → Treatment A

Marker Level (+) → Randomize → Treatment B

Power trial separately within marker groups

Sargent et al., JCO 2005

Subset Analysis – when overall treatment effect is not significant

- Split the overall Type I error rate of $\alpha = 0.05$
  - $\alpha_1$ for the primary comparison of treatment to control in all patients
  - $0.05 - \alpha_1$ for treatment to control in marker defined subgroup, when overall comparison is not significant at $\alpha_1$

- Appropriate when:
  - New treatment may be broadly effective
  - Marker (and threshold) known at outset of trial
  - Subset analysis is a secondary fall back option
Adaptive Designs
Jiang et al. – 2007: Biomarker adaptive threshold design

- **Marker known, but threshold unknown for defining positive/negative**
  - Incorporate a statistically valid hypothesis for identification and validation of a cut point for a pre-specified single biomarker (or composite score)

- **Procedure:**
  - Split the overall Type I error rate
  - Compare treatment outcomes on all pts at $\alpha_1$
  - If not significant, perform second test at $0.05 - \alpha_1$ in subset identified by a marker cut point that results in the largest treatment effect
  - p-value by permutation test

Adaptive Designs
Freidlin, Jiang, Simon CCR 2010: Cross validated adaptive signature design

- **Signature and threshold, both unknown**
- Test for overall effect at level $\alpha_i < \alpha$
- If no overall effect, perform K-fold cross validation for signature development and subset effect testing -
  - Split sample into N-M development and M validation sets,
  - Develop signature in development set,
  - Apply to identify sensitive patients in validation set,
  - Repeat K times
- At the end of this CV process, each patient is classified as sensitive or not-sensitive
Adaptive Designs
Freidlin, Jiang, Simon CCR 2010: Cross validated adaptive signature design

- Compare outcomes between the experimental and standard therapy for the sensitive patients (identified from the CV).

- If significant treatment effect, then:
  - Final classifier is obtained by applying the signature development procedure to the entire study population.
  - The (conservative) treatment effect for the sensitive subset of future patients is obtained by the CV method (p-value based on permutation test).

So, what dictates sample size in biomarker driven trials?

- Purported treatment effect in the sensitive subset, and in the overall population
- Prevalence of the marker
- Accuracy of the assay / signature used to classify patients
- All of the above dictates the optimal design, which ultimately defines sample size.
Conclusions

- Choice of trial design depends on clinical, statistical and ethical considerations.

- Relevant:
  - marker prevalence,
  - incremental benefit of marker-based selection,
  - strength of preliminary data

Recommendations

- If possible, do not restrict population; prospectively design trial to address biomarker hypothesis

- As knowledge of biology increases, enrichment designs will become more popular

- Phase II setting- Employ adaptive strategies, when reliable short term endpoint is available
Thank You