Plenary Session 3: Phases of Clinical Development of Therapeutics: Phase III

Statistics for Clinical Trials Part II: Interim Trial Analysis, Comparisons and Analysis of Correlative Studies

Chris O’Callaghan

(Dongsheng Tu*)
5 statisticians and 5 epidemiologists are travelling together on a train. They all start chatting and it transpires that all the epidemiologists have bought a ticket, but the statisticians have only bought 1 between the 5 of them. “Why did you do that?” asks one of the epidemiologists. “Surely you’re going to get caught and thrown off the train?”

“Just wait and see!” says one of the statisticians.

As the ticket inspector is approaching to check everyone’s tickets, the statisticians all go off to the nearest toilet – the inspector passes the epidemiologists and inspects all their tickets then moves on and notices that the toilet is locked. “Tickets please!”, shouts the inspector. One of the statisticians pushes their ticket under the toilet door, which the inspector checks and returns under the door. Once the inspector has gone, all the statisticians return to their seats to the awe and amazement of the epidemiologists. “That’s incredibly clever!” says one of the epidemiologists.

A few weeks later they all find themselves on the same train again. They sit together and start chatting once more. “We’ve done what you suggested”, says one of the epidemiologists. “And just bought one ticket between the five of us!” “Oh really”, says one of the statisticians. “we haven’t bought ANY tickets this time!” The epidemiologists look at each other in amazement. “OK, one ticket between you is fine but not buying any at all is ludicrous!” says the statistician. “We have our methods!”, smiles one of the statisticians.

As the ticket inspector approaches they hurry off to the toilet. Once they’re inside, the statisticians follow them. “Tickets please!” shouts one of the statisticians.
Topics to be covered

• Interim Analyses
  – Simulation – Multiple Analyses
  – Group Sequential Testing
  – Negative Stopping
  – Examples**

• Analysis of Correlative Studies
  – Prognostic Markers
  – Predictive Markers
  – Statistical differentiation of the two
  – Examples**

** - extra examples provided
Interim Analyses

• For ethical reasons, it is often desirable to examine the efficacy results of a trial before it is complete
  – Usually this is because of a concern that one arm may already be demonstrably superior, but sometimes the issue is the futility of demonstrating a difference
  – Can we:
    • Reduce the number of patients randomized?
    • Reduce the risk of adverse events to patients?
    • Offer patients the superior therapy?
Interim Analyses

- One, two or even more interim analyses may be considered depending on the sample size, duration and outcome of the trial.

- However, repeated testing results in accumulating type I error...the chance you will conclude there is a benefit when in reality there is not = "false benefit"
Experimental Errors

State of Nature (Reality)

<table>
<thead>
<tr>
<th>Results of Statistical Analysis</th>
<th>State of Nature (Reality)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Effect</td>
<td>No Effect</td>
</tr>
<tr>
<td></td>
<td>'Accept' null hypothesis when it is true</td>
</tr>
<tr>
<td>Effect</td>
<td>Type I</td>
</tr>
<tr>
<td></td>
<td>((\alpha, \beta)) error</td>
</tr>
<tr>
<td></td>
<td>Reject null hypothesis when it is true</td>
</tr>
<tr>
<td></td>
<td>&quot;consumers/regulatory risk&quot;</td>
</tr>
<tr>
<td>Effect</td>
<td>Type II *</td>
</tr>
<tr>
<td></td>
<td>((\beta)) error</td>
</tr>
<tr>
<td></td>
<td>'Accept' null hypothesis when it is false</td>
</tr>
<tr>
<td></td>
<td>&quot;sponsors risk&quot;</td>
</tr>
</tbody>
</table>

* Power = 1-type II error = Probability of correctly rejecting \(H_0\) (probability of rejecting the null hypothesis given that the alternate hypothesis is true)
The Problem of Multiple Testing

- Roll a dice. 1 in 6 chance of a six = 17%
- Roll it 3 times and what is the chance that you will get at least one six?
  - $1 \& 1 \& 1, 1 \& 1 \& 2, 1 \& 1 \& 3, \ldots 6 \& 6 \& 6$
  - $1 - (\text{the chance of NOT getting a six on any of the three roll})$
  - $1 - (5/6)(5/6)(5/6) = 42\%$
- Roll it 10 times and what is the chance that you will get at least one six?
  - $1 - (5/6)^{10} = 84\%$
Anyone who cannot cope with mathematics is not fully human. At best he is a tolerable subhuman, who has learned to wear shoes, bathe, and not make messes in the house.

Robert Heinlein
The Problem of Multiple Analyses

- Conduct a trial where the reality is no difference between the arms and where you accept a pre-specified 5% chance of an erroneous result (declaring there to be a difference when the reality is there is not one).

- Conduct the trial 2 times and what is the chance that you will conclude there is a difference at least once?
  \[= 1 - (1 - 0.05)^2 = 1 - (0.95)^2 = 9.8\%\]

- Conduct the trial 8 times and what is the chance that you will conclude there is a difference at least once?
  \[= 1 - (1 - 0.05)^8 = 1 - (0.95)^8 = 34\%\]

- Inflation of the Experimentwise Error Rate (False Discovery Rate)
The Problem of Multiple Analyses - Simulation

- 3-year accrual period, and a final analysis one year later
- 60 patients on each arm
- Lifetimes follow same distribution (exponential distribution with a median survival of 1 year)
- In reality, no difference in survival between the two groups
- This simulation is repeated 100 times
The Problem of Multiple Analyses - Simulation

- 5 situations considered
  - 1 logrank test – at conclusion (4 years)
  - 2 logrank tests – every 2 years
  - 4 logrank tests – every year
  - 8 logrank tests – every 6 months
  - 16 logrank tests – every 3 months
The Problem of Multiple Analyses - Simulation

- Logrank p value was <0.05 at:
  - the final test (4 years) in $5 \text{ of } 100$
  - either the 2 or 4 year test in $10 \text{ of } 100^{**}$
  - at least 1 of the 4 yearly tests in $17 \text{ of } 100$
  - at least 1 of 8 semiannual tests in $21 \text{ of } 100$
  - at least 1 of 16 3-month tests in $26 \text{ of } 100$
The Problem of Multiple Analyses - Simulation

- Risk of analyzing the data at a “random high”
- ** 2 & 4 year p-values for the 10 ‘single interim-analysis’ studies with a p<0.05:

<table>
<thead>
<tr>
<th>At the 4 year analysis</th>
<th>At the 2 year analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>p values at</td>
<td>p values at</td>
</tr>
<tr>
<td>2 years</td>
<td>2 years</td>
</tr>
<tr>
<td>4 years</td>
<td>4 years</td>
</tr>
<tr>
<td>0.1194</td>
<td>0.0220</td>
</tr>
<tr>
<td>0.4417</td>
<td>0.0205</td>
</tr>
<tr>
<td>0.7104</td>
<td>0.0165</td>
</tr>
<tr>
<td>0.3704</td>
<td>0.0086</td>
</tr>
<tr>
<td>0.0734</td>
<td>0.0110</td>
</tr>
</tbody>
</table>

4 years 2 years 4 years 2 years
Interim Analyses

- This is not just a theoretical problem
  - Examination of practices of trials groups indicates many studies were stopped “too soon” when interim analyses were repeatedly conducted and reported
  - Montori *et al.* conducted a systematic review
    - 143 RCTs stopped early for experimental benefit
    - 92 published in 5 high-impact journals
    - From 0.5% of all RCTs published in 1990-1994 to 1.2% in 2000-2004 (*p*<0.001 for trend)
**Interim Analyses**


- “This survey assessed the level of attention to the problem of multiple comparisons in the analyses of contemporary randomized clinical trials.

- Of the 67 trials surveyed, 66 (99 percent) performed multiple comparisons with a mean of 30 therapeutic comparisons per trial.

- When criteria for statistical impairment were applied, 50 trials (75 percent) had the statistical significance of at least one comparison impaired by the problem of multiple comparisons, and 15 (22 percent) had the statistical significance of all comparisons impaired by the problem of multiple comparisons.”
CALGB 9633
Adjuvant Chemotherapy in Stage IB NSCLC

• ASCO 2004
  - Median follow-up = 34 months... final analysis planned at 150 deaths
  - 36 deaths (/173) in chemo arm vs 52 (/171) in obs. arm (88 deaths)
  - Overall Survival HR=0.62; 95% CI: 0.41-0.95, \( p=0.028 \)
  - (Slow accruing) study closed early by DSMB as \( p \) value for OS less than a prespecified stopping boundary

• ASCO 2006
  - Median follow-up = 52 months, 131 deaths had occurred
  - Overall Survival HR = 0.80; 90% CI = 0.60-1.07, \( p=0.10 \)
  - “Final” analysis still to be conducted at 150 deaths
  - Conclusion? = Now underpowered for small differences
Current Practice

- Group Sequential Designs by far the most prevalent approach
  - Data are analyzed in groups when a pre-specified amount of information (e.g., 25%, 33%, 50% of the events) is available
  - The critical value of the tests (or the significance level) at each interim analysis is adjusted for multiple comparisons so the overall type I error is less than the nominal level
Group Sequential Tests

- Pocock (1977)  
  - divides equally the overall significance levels

- Peto (1976)  
  - interim analyses with .001 nominal level so that the final analysis is closed to .05

- O’Brien (1979)  
  - started with stringent nominal levels and gradually increased to a level close to .05

- Fleming (1984)  
  - with less extreme early nominal level
### Nominal Significance Levels for 2-sided 5-stage Group Sequential Trials Maintaining Overall Significance Level of 0.05

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.016</td>
<td>0.001</td>
<td>0.00001</td>
<td>0.0051</td>
</tr>
<tr>
<td>0.016</td>
<td>0.001</td>
<td>0.0013</td>
<td>0.0061</td>
</tr>
<tr>
<td>0.016</td>
<td>0.001</td>
<td>0.008</td>
<td>0.0073</td>
</tr>
<tr>
<td>0.016</td>
<td>0.001</td>
<td>0.023</td>
<td>0.0089</td>
</tr>
<tr>
<td>0.016</td>
<td>0.049</td>
<td>0.041</td>
<td>0.0402</td>
</tr>
</tbody>
</table>
Stopping Boundaries

O'Brien and Fleming Stopping Boundaries

Pocock Stopping Boundaries
Lan & DeMets (1983)

- When the number of interim analyses is not fixed and the time of analysis is not pre-specified
- Lan & Demets proposed a stopping boundary which is a function of past and current but not future decision-times
- Define alpha-spending function
  - governs the rate at which the overall $\alpha$ is to be spent
# events = 160; 46% of target of 350 events
Stopping when the experimental arm does not appear to help (futility)

- Assume that $P_A$ and $P_B$ are response rates respectively for control and experimental arm
  - Perform an analysis when we have about half the sample size
  - stop if observed response rate for the study treatment is lower than that of the control
- This leads to reduction of expected sample size if the test treatment is ineffective
- For time to an event outcome, perform the analysis when half of required number of events are observed and stop if observed hazard ratio (B to A) equals or exceeds 1.
- This may or may not lead to a reduction of sample size
Example of Interim Analysis
Plan in the protocol

“We are planning two interim analyses to allow early termination of the study if the results are extreme. After observing one third and two thirds of the expected recurrences from the disease-free survival analysis, i.e., 174 and 348 recurrences respectively, we will perform a log-rank test on the primary endpoint using the O’Brien-Fleming type boundaries as proposed by Lan and Demets. We expect to have 174 recurrences approximately half a year after the end of accrual and 348 recurrences approximately 2.2 years after the end of accrual.

The results of the interim analyses will be presented to the monitoring committee. Early termination will be considered when a significance level of the first and second interim analyses are less than 0.0004 and 0.0129 respectively. The nominal significance value for the final analysis is 0.0457. This group sequential procedure is based on the type I error spending function as proposed by Lan and Demets such that the overall significance level will be maintained at 5%.”
Data & Safety Monitoring Committee

- Membership composed of physicians, statisticians, other scientists, lay representatives

- Responsibilities include review of interim analyses of outcome data and cumulative toxicity data summaries, trial performance information such as accrual information, reports of related studies both internal and external to the group and major modifications proposed to the study.
Statistical thinking will one day be as necessary a qualification for efficient citizenship as the ability to read and write.

H.G. Wells
Cancer Treatment and Biomarkers

- Many drugs are found to improve disease free or overall survival for patients with various types of cancer.
- However, no regimen is found universally effective for all patients.
- The selection of a particular treatment which is best for a given patient is challenging and currently more of an art than a science.
- There is a need to find good biomarkers which would be used to “personalize” treatment for cancer patients.
Cancer Treatment and Biomarkers

• Many drugs are found to improve disease-free or overall survival for patients with various types of cancer.

• However, no regimen is found universally effective for all patients.

• The selection of a particular treatment which is best for a given patient is challenging and currently more of an art than a science.

• There is a need to find good biomarkers which would be used to “personalize” treatment for cancer patients.
Types of Tumor Biomarkers

- Prognostic markers
- Predictive markers
Prognostic markers

- The biomarker is called prognostic if it provides information concerning the anticipated natural history of the disease process in a given individual.
- ...but where the outcome is independent from therapy.
- Answers the question "When?"
- Example: Prostate specific antigen (PSA) in prostate cancer which is used to classify the risk of the patients.
Predictive biomarkers

• A predictive marker is a marker that allows the prospective identification of individuals who will or will not benefit from the use of a particular therapy.

• Predicts the outcome of a specific therapy.

• Answers question “With what?” or “How much?”

• Example: Estrogen receptor in breast cancer which is used to select hormonal treatments for the breast cancer.
- **Differential Efficacy**
- Parallel *versus* non-parallel lines
- In statistical terms this is termed **interaction** and can be specifically tested for, i.e. a p-value for interaction can be generated.
- Assuming there is sufficient power, this can be used to assess the null hypothesis that there is no differential efficacy between the therapies (no interaction) or that the marker is not predictive of efficacy.
Example: K-ras as a Biomarker in Colorectal Cancer

The NEW ENGLAND JOURNAL of MEDICINE

K-ras Mutations and Benefit from Cetuximab in Advanced Colorectal Cancer

Christos S. Karapetis, M.D., Shirin Khambata-Ford, Ph.D., Derek J. Jonker, M.D., Chris J. O’Callaghan, Ph.D., Dongsheng Tu, Ph.D., Niall C. Tebbutt, Ph.D., R. John Simes, M.D., Haji Chalchal, M.D., Jeremy D. Shapiro, M.D., Sonia Robitaille, M.Sc., Timothy J. Price, M.D., Lois Shepherd, M.D.C.M., Heather-Jane Au, M.D., Christiane Langer, M.D., Malcolm J. Moore, M.D., and John R. Zalcberg, M.D., Ph.D.*
The Influence of *K-ras* Exon 2 Mutations on Outcomes in

A Randomized Phase III Trial of Cetuximab + Best Supportive Care (BSC) versus BSC Alone in Patients with Pre-treated Metastatic EGFR-Positive Colorectal Cancer (NCIC CTG CO.17)

A trial of the

National Cancer Institute of Canada Clinical Trials Group (NCIC CTG)

and the

Australasian Gastro-Intestinal Trials Group (AGITG)
Cetuximab: Multiple Mechanisms of Action

- IgG1 monoclonal antibody
- Binds to EGFR and competitively inhibits ligand binding (e.g. EGF)
- Blocks receptor dimerization, tyrosine kinase phosphorylation, and signal transduction
- IgG1-induced Antibody-Dependent Cell Cytotoxicity (ADCC)

# Cetuximab: Phase II Clinical Data

<table>
<thead>
<tr>
<th>Study</th>
<th>Treatment</th>
<th>N</th>
<th>ORR</th>
<th>TTP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Irinotecan Failure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saltz L. (J\ Clin Oncol\ 2004 (IMC 0141))</td>
<td>Cetuximab</td>
<td>57</td>
<td>8.8%</td>
<td>1.4 mo</td>
</tr>
<tr>
<td>Cunningham D. (N Engl J Med\ 2004 (EMR 007 / BOND))</td>
<td>Cetuximab</td>
<td>111</td>
<td>10.8%</td>
<td>1.5 mo</td>
</tr>
<tr>
<td></td>
<td>Cetuximab + Irinotecan</td>
<td>218</td>
<td>22.9%</td>
<td>4.1 mo</td>
</tr>
<tr>
<td><strong>Irinotecan, Oxaliplatin, Fluoropyrimidine Failure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lenz H-J. (J Clin Oncol\ 2006 (IMC 0144))</td>
<td>Cetuximab</td>
<td>346</td>
<td>12.4%</td>
<td>1.4 mo</td>
</tr>
</tbody>
</table>
NCI C CTG CO.17: Randomized Phase III Trial in mCRC

Failed or intolerant to all recommended therapies, ECOG 0-2, No Prior EGFR directed therapy

Primary Endpoint:

Secondary Endpoints:

Overall Survival
Progression Free Survival
Objective Response Rate (RECIST criteria)
Safety and Quality of Life

Cetuximab* + BSC

BSC alone

* Cetuximab 400 mg/m² IV week 1 then 250 mg/m² IV weekly
NCI C CTG CO.17: Accrual

Weeks Post Central Activation

20 months

Final
n=572

n=320

Target Accrual

n=252
NCI C CTG CO.17: Subject Disposition

Registered
N = 1243*

EGFR detectable; N = 981 (79%)

Randomized
N = 572

Cetuximab
N = 287

BSC
N = 285

No Cetuximab
N = 4

Treated
N = 288

N = 5

Treated
N = 274

Withdrawn Consent
N = 6

Prior to Progression
N = 15

Post Progression
N = 274

On Treatment
N = 17

Off Treatment
N = 271

• Deaths (N = 12)
• PD (N = 205)
• Symptomatic progression (N = 27)
• Drug toxicity (N = 9)
• Subject request (N = 10)

* Patients were allowed to be enrolled at the time of previous chemotherapy
### NCI C CTG CO.17: Overall Survival

<table>
<thead>
<tr>
<th>Study arm</th>
<th>MS (months)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cetuximab + BSC</td>
<td>6.1</td>
<td>5.4 - 6.7</td>
</tr>
<tr>
<td>BSC alone</td>
<td>4.6</td>
<td>4.2 - 4.9</td>
</tr>
</tbody>
</table>

HR 0.77 (95% CI = 0.64 – 0.92)

Stratified log rank
p-value = 0.0046

---

**Subjects at Risk**

<table>
<thead>
<tr>
<th>Study arm</th>
<th>Subjects at Risk</th>
<th>Months</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CET+BSC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>287</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>217</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>136</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>78</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>BSC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>285</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>197</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>85</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>44</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>27</td>
</tr>
</tbody>
</table>

---

*Jonker et al., NEJM 2007*
NCI C CTG CO.17: Progression Free Survival

### Study arm | Med PFS (months) | 95% CI
---|---|---
Cetuximab + BSC | 1.9 | 1.8 - 2.1
BSC alone | 1.8 | 1.8 - 1.9

HR 0.68 (95% CI =0.57 – 0.80)

Stratified log rank p-value < 0.0001
Which patients benefit?

A reliable biomarker is needed:

- to provide an accurate prediction of who will respond and benefit from cetuximab
- to improve the therapeutic index
- to improve cost effectiveness of EGFR monoclonal antibody based therapy of pre-treated colorectal cancer

The predictive value of the biomarker would need to be differentiated from its prognostic implications

The *K-ras* mutation status of the bowel cancer may be such a marker of response and a predictor of benefit.
EGFR Signaling Cascade and K-ras

K-ras is a small G protein
Self inactivating - from GDP to GTP state
Switched off by intrinsic GTPase activity
K-ras mutation leads to constitutive activation mediated through reduced GTPase activity
Inhibitors upstream may be ineffective
**KRAS Mutation Detection**

- DNA extracted from slides containing FFPE tissue sections
- *KRAS* exon 2 is amplified by PCR and subjected to bidirectional sequencing
- Sequence traces are analyzed by mutation detection software & visual inspection
- Mutations are most common on codons 12 & 13

Wild Type

![Wild Type Sequence](image1)

Mutant

![Mutant Sequence](image2)
<table>
<thead>
<tr>
<th>Reference</th>
<th>Treatment</th>
<th>Number</th>
<th>ORR %</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Lievre, A et al</em></td>
<td>Cetuximab +/- CT</td>
<td>89</td>
<td>40</td>
</tr>
<tr>
<td><em>Di Fiore, F et al</em></td>
<td>Cetuximab + CT</td>
<td>59</td>
<td>28</td>
</tr>
<tr>
<td><em>Khambata-Ford et al</em></td>
<td>Cetuximab</td>
<td>80</td>
<td>10</td>
</tr>
<tr>
<td><em>De Roock, W et al</em></td>
<td>Cetuximab +/- CT</td>
<td>108</td>
<td>41</td>
</tr>
</tbody>
</table>
NCI C CTG CO.17  *K-Ras* Analysis

- No difference between *K-ras* mutated and WT patients re: demographics, previous treatment or other variables

N=572 randomized: ITT subset

N=394: *K-ras* assessed subset (69%)

- N=164 (42%) mutant
- N=230 (58%) wild-type
## Comparison of ITT and K-ras assessed subsets

<table>
<thead>
<tr>
<th>Baseline Characteristic</th>
<th>ITT (N = 572)</th>
<th>Mutated K-ras (N = 164)</th>
<th>Wild-type K-ras (N = 230)</th>
<th>p-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age - median</td>
<td>63.2</td>
<td>62.0</td>
<td>63.5</td>
<td>0.569</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>204 (35.7)</td>
<td>63 (38.4)</td>
<td>74 (32.2)</td>
<td>0.200</td>
</tr>
<tr>
<td>M</td>
<td>368 (64.3)</td>
<td>101 (61.6)</td>
<td>156 (67.8)</td>
<td></td>
</tr>
<tr>
<td>ECOG PS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>136 (23.8)</td>
<td>34 (20.7)</td>
<td>56 (24.3)</td>
<td>0.695</td>
</tr>
<tr>
<td>1</td>
<td>302 (52.8)</td>
<td>94 (57.3)</td>
<td>127 (55.2)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>134 (23.4)</td>
<td>36 (22.0)</td>
<td>47 (20.4)</td>
<td></td>
</tr>
<tr>
<td>Prior XRT</td>
<td>202 (35.3)</td>
<td>50 (30.5)</td>
<td>77 (33.5)</td>
<td>0.531</td>
</tr>
<tr>
<td>Prior chemoRx</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>adjuvant</td>
<td>211 (36.9)</td>
<td>57 (34.8)</td>
<td>83 (36.1)</td>
<td>0.786</td>
</tr>
<tr>
<td>antiTS</td>
<td>572 (100.0)</td>
<td>164 (100.0)</td>
<td>230 (100.0)</td>
<td></td>
</tr>
<tr>
<td>irinotecan</td>
<td>550 (96.2)</td>
<td>161 (98.2)</td>
<td>219 (95.2)</td>
<td>0.119</td>
</tr>
<tr>
<td>oxaliplatin</td>
<td>559 (97.7)</td>
<td>163 (99.4)</td>
<td>222 (96.5)</td>
<td>0.060</td>
</tr>
<tr>
<td>Arm</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CET</td>
<td>287 (50.2)</td>
<td>81 (49.4)</td>
<td>117 (50.9)</td>
<td>0.772</td>
</tr>
<tr>
<td>BSC</td>
<td>285 (49.8)</td>
<td>83 (50.6)</td>
<td>113 (49.1)</td>
<td></td>
</tr>
</tbody>
</table>

*p-value* between mutated and wild-type K-RAS groups from chi-square test for categorical variables and t-test for continuous variables.
NCI C CTG C0.17: Primary endpoint overall survival

Total study population (ITT analysis)  K-ras assessed subset
NCI C CTG C0.17: PFS in the Mutant *K-ras* Subgroup

<table>
<thead>
<tr>
<th>Study arm</th>
<th>Med PFS (months)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cetuximab + BSC</td>
<td>1.8</td>
<td>1.7 - 1.8</td>
</tr>
<tr>
<td>BSC alone</td>
<td>1.8</td>
<td>1.7 - 1.8</td>
</tr>
</tbody>
</table>

HR 0.99, 95% CI (0.73, 1.35)

Log rank p-value: 0.96
**NCI C CTG C0.17: PFS in the *K-ras* Wild-Type Patients**

<table>
<thead>
<tr>
<th>Study arm</th>
<th>Med PFS (months)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cetuximab + BSC</td>
<td>3.8</td>
<td>3.1 - 5.1</td>
</tr>
<tr>
<td>BSC alone</td>
<td>1.9</td>
<td>1.8 - 2.0</td>
</tr>
</tbody>
</table>

**HR 0.40 95% CI (0.30, 0.54)**

Log rank p-value: **<0.0001**

**Test for Interaction**

p < 0.001
NCI C CTG C0.17: Overall survival in *K-ras* Mutant patients

**Proportion Alive**

- **Cetuximab**
- **BSC**

**Study arm**

<table>
<thead>
<tr>
<th>Study arm</th>
<th>MS (months)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cetuximab + BSC</td>
<td>4.5</td>
<td>3.8 - 5.6</td>
</tr>
<tr>
<td>BSC alone</td>
<td>4.6</td>
<td>3.6 - 5.5</td>
</tr>
</tbody>
</table>

**HR 0.98  95% CI (0.70,1.37)**

**Log rank p-value: 0.89**
NCI C CTG C0.17: Overall survival in *K-ras* Wild-Type patients

### Survival Analysis

**Study arm** | **MS (months)** | **95% CI**
--- | --- | ---
Cetuximab + BSC | 9.5 | 7.7 - 10.3
BSC alone | 4.8 | 4.2 - 5.5

**HR** 0.55 95% CI (0.41, 0.74)

**Log rank p-value:** <0.0001

**Test for Interaction**

\[ p = 0.01 \]

---

<table>
<thead>
<tr>
<th>Study arm</th>
<th>MS (months)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cetuximab + BSC</td>
<td>9.5</td>
<td>7.7 - 10.3</td>
</tr>
<tr>
<td>BSC alone</td>
<td>4.8</td>
<td>4.2 - 5.5</td>
</tr>
</tbody>
</table>

---

**Log rank p-value:** <0.0001

**Test for Interaction**

\[ p = 0.01 \]
NCI C CTG C0.17: Overall Survival by *K-ras* Status in BSC ARM

![Graph showing overall survival by K-ras status with Kaplan-Meier curves.](image)

<table>
<thead>
<tr>
<th>KRAS status</th>
<th>MS (months)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mutated</td>
<td>4.6</td>
<td>3.6 - 5.5</td>
</tr>
<tr>
<td>Wild-Type</td>
<td>4.8</td>
<td>4.2 - 5.5</td>
</tr>
</tbody>
</table>

HR 1.01 95% CI (0.74, 1.37)

Log rank p-value: 0.97

**NO PROGNOSTIC IMPACT OF K-ras STATUS**
NCIC CTG CO.17: K-ras and Cetuximab Conclusions

In the context of pre-treated advanced colorectal cancer:

- There is no benefit in using cetuximab monotherapy in patients that have mutated K-ras tumours.
- There is a 4.7 month improvement in median survival with cetuximab in patients with K-ras wild-type tumours.
- The p-value for the interaction between K-ras status and treatment is 0.01.
- There is an improvement in PFS with cetuximab in K-ras wild-type tumours.
- K-ras mutation status does not have a treatment-independent prognostic effect.
NCI C CTG CO.17: Additional Correlative Studies

• Approved
  - Epiregulin & Amphiregulin expression – ASCO 2009
  - BRAF mutations, PIK3CA mutations, Loss of PTEN (IHC, FISH) – in progress
  - K-Ras validation – pending FDA/BMS

• Proposed
  - FCγR polymorphisms
  - IGF-1R expression
“Play some Frisbee, chew on an old sock, bark at a squirrel. If that doesn’t make you feel better, eat some cheese with a pill in it.”
Interim Analyses

Examples
Example I: Toxic Deaths?
(NCI C CTG BR.8)

- To determine whether the CODE regimen plus thoracic irradiation is superior to standard alternating CAV/EP (Murray et al. 1998)
  - Activated in July 1992
  - Planned sample size = 410 over 2.5 years + 8 months of follow-up to realize 280 events (HR 1.4, 2-sided alpha=5%, Power = 80%)
  - Interim analysis initially planned at 100 events (36%) with early stopping for benefit if p<0.0012
  - No futility analysis boundary specified
  - 109 and 110 eligible patients in CAV/EP and CODE arms respectively at time of interim analysis in April 1996 (4 years post activation)
Causes of Death

<table>
<thead>
<tr>
<th>Cause</th>
<th>CAV/EP</th>
<th>CODE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease</td>
<td>88</td>
<td>73</td>
</tr>
<tr>
<td>Protocol Treatment Complication</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>Disease and Non Protocol Treatment Treatment</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Other Causes</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Cause Unknown</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

- Excessive deaths due to toxicity in the CODE arm?
### Main Results of Analyses to DSMC

#### Univariante Analysis on Treatment Effect

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Coeff</th>
<th>Stderr</th>
<th>P-value</th>
<th>RR/OR</th>
<th>95% CI for RR/OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survival(^1)</td>
<td>0.0604</td>
<td>0.1489</td>
<td>0.6851</td>
<td>1.06</td>
<td>(0.79, 1.42)</td>
</tr>
<tr>
<td>Progression(^1)</td>
<td>0.1536</td>
<td>0.1559</td>
<td>0.3246</td>
<td>1.17</td>
<td>(0.86, 1.58)</td>
</tr>
<tr>
<td>Time-to-response(^1)</td>
<td>-0.2133</td>
<td>0.3019</td>
<td>0.4799</td>
<td>0.81</td>
<td>(0.15, 1.46)</td>
</tr>
<tr>
<td>Toxic deaths(^2)</td>
<td>2.1366</td>
<td>1.0695</td>
<td>0.0457</td>
<td>8.47</td>
<td>(1.04, 68.9)</td>
</tr>
</tbody>
</table>

**Note:**  
RR indicates the ratio of the hazards of CAV/EP divided by that of CODE  
1 These analyses were done using simple Cox regression model  
2 Logistic regression model using deaths due to treatment as event  
RR/OR for logistic model indicates an odds ratio
DSMC Actions and Decisions

• The decision of the DSMC at the time of interim analysis was to monitor toxic deaths closely and continue.

• The DSMC recommended termination after an additional conference call for the DSMC members one month after the interim analysis.

• An expedited report from the DSMC chair was sent to the NCIC CTG central office on the same day.

• The Clinical Trial Committee accepted the DSMC recommendation of terminating the study.
Example II: Inferior experimental arm? (NCIC CTG PA.1)

- To determine whether BAY 12-9566 improves overall survival as compared to gemcitabine in patients with unresected locally advanced or metastatic adenocarcinoma of pancreas

- Activated in Dec. 1997

- Two planned interim analyses
  - First based on PFS
  - Second based on OS
First Interim Analysis

- When 30 patients were accrued in each arm and followed for at least 8 weeks
- Based on the 8-week progression free rate (PFR)
  - Study would be stopped when 6 or less out of 30 patients (20%) in the test arm were free of progression at 8 weeks
  - 97.4% chance to stop the study and conclude the BAY is inactive when the actual 8-week PFR is 10% and 84% chance to continue the study when the actual 8-week PFR is 30%
- 11 (31%) patients on BAY arm
  - 16 (50%) on Gem were free from the progression
  - DSMC recommended continuation
Second Interim Analysis

- Conducted when 140 deaths were observed
- 138 and 139 eligible patients in BAY and GEM arms respectively
- Based on overall survival
  - Study would be stopped if the p-value of 2-sided log-rank test was less than 0.0056 based on O’Brien and Fleming boundary
- Median survivals in the analysis
  - Gem 6.4 months vs. BAY 3.2 months [p=0.0001]

DSMC recommended study closure
Overall Survival in PA.1

![Graph showing overall survival in PA.1 with time in months and percentage remaining. The graph compares Bay 12-9566 and Gemcitabine with # At Risk values at various time points.](image)
Example III: Not superior experimental arm? (NCIC CTG MA.19)

- To determine whether DPPE+DOX improves progression-free survival (PFS) as compared to DOX alone in patients with metastatic/recurrent breast cancer.

- Activated in Feb. 1998 with planned sample size of 350 to be accrued over 2 years with additional 1 year of follow-up to realize 256 progressions (6 month to 9 month median PFS, 2-sided alpha=5%, Power=90%).

- Single interim analysis planned when first 150 patients accrued had been followed for at least 3 months, but including assessments of both:
  - Response Rate
  - PFS
Interim Analysis

• When first 150 patients accrued had been followed for at least 3 months

• Based on the response rate (RR)
  – Study would be stopped if the observed RR on DPPE/DOX arm is not superior to that on DOX alone arm by more than 5%
  – More than 90% chance to continue the study when the actual RR for DPPE/DOX and DOX alone were respectively $\geq 45\%$ and $\leq 30\%$

• Observed RR in the analysis: 35.5% for DPPE/DOX and 36.5% on DOX alone

DSMC recommended trial be stopped but suggested final analysis be performed according to protocol
Final Analysis

- **Response rate**
  - DPPE/DOX: 28.8% vs. DOX alone: 29.0% (p=0.95)

- **Median Progression Free Survival**
  - DPPE/DOX: 5.9 months vs. DOX alone: 6.0 months (p=0.31)

- **Medium Overall Survival**
  - DPPE/DOX: 23.6 months vs. DOX alone: 15.6 months (p=0.021)
Overall survival in MA.19
Example IV: Superior experimental arm? (NCIC CTG SR.2)

- To determine if there is a difference in the incidence of wound healing complications in patients with extremity soft tissue sarcoma treated by pre- or post-operative external beam radiotherapy.

- Activated in Oct. 1994 with a planned sample size of 266 over 5 years and assuming 30% complication rate in pre-op arm and powered to 80% to detect an absolute decrease of 15% in wound complication rate in post-op arm with 2-sided alpha of 5%.

- Single planned interim analysis.
Interim Analysis

- When first 133 eligible patients were accrued and evaluated

- Based on the wound complication rate (WCR)
  - Study would be stopped if the p-value of 2-sided Fisher’s exact test was less than 0.0056 based on O’Brien and Fleming boundary

- Observed WCRs in the analysis: 36% for pre-op arm and 14% for post-op (p=0.0050)

DSMC recommended stopping the trial based on the p-value for WCR or redesign of the trial using overall survival as the primary endpoint
Example V:
(NSABP B-14 - ReRandomization)

• Breast cancer patients with estrogen receptor-positive tumours and no evidence of axillary node involvement who had completed 5 years of tamoxifen, free of recurrence or other events were randomized to:
  – A: tamoxifen for an additional 5 years
  – B: placebo
Background

• **Primary Endpoint (DFS)**
  - time to either breast cancer recurrence at a local, regional, or distant anatomic site
  - the occurrence of a contralateral breast cancer or other primary malignancy
  - death from any cause

• **Sample Size**
  - to detect a relative 40% reduction from a 5% failure rate in the placebo arm at two-sided 10% type I error required 115 events
Planned Interim Analyses

- Beginning in the 4th year at about 1 to 1.5 year intervals
  - corresponding to equal increments of the requisite events

- Fleming *et al.* early stopping rule:
  - 5 two-sided 10% stopping boundaries
    - 0.00244, 0.00302, 0.00346, 0.00434, 0.09761

- First interim analysis was unremarkable
Second Interim Analysis

- **Number of events**
  - tamoxifen arm 43/587
  - placebo arm 24/573
  - $p = 0.028$
  - Early stopping criterion = 0.00244

- **Number of deaths**
  - tamoxifen arm 19/587
  - placebo arm 10/573

DSMC was concerned, but recommended continuation
Figure 1  NSABP B-14: Disease-free survival comparison at second interim analysis.
Third Interim Analysis

- **Number of events**
  - tamoxifen arm 56/591
  - placebo arm 32/575
  - \( p = 0.015 \)
  - Early stopping criterion = 0.00346

- **Number of deaths**
  - tamoxifen arm 23/591
  - placebo arm 13/575

Despite the fact that the early stopping criterion was not crossed, the DSMC recommended stopping the study.
Figure 2  NSABP B-14: Disease-free survival comparison at third interim analysis.
• Based on Rule for stopping the study when experiment arm doesn’t appear to help
  - perform Interim analysis when one half of the required events has taken place
  - stop if the risk ratio for the standard arm over the experimental arm is less than 1.0
  - We would have stopped at the 2\textsuperscript{nd} interim analysis with a loss of power < 0.02 for any alternative hypothesis indicating a treatment benefit

• The estimated hazards ratio at 3rd interim analysis was 0.59
  - 95\% CI 0.38-0.90
  - Conclusion: no additional benefit for continued tamoxifen
Example VI: Is an interim analysis needed?

(NCI C CTG CO.20)

• Does the addition of Brivanib to Cetuximab improve overall survival in patients with end-stage metastatic colorectal cancer who have failed all other standard chemotherapy?

• Double-blind, placebo controlled trial. Sample size of 750 to be accrued over 2 years (approx. 30 patients per month) with an additional 3 months of follow-up to realize 527 deaths, necessary to provide 90% power to detect a HR=0.75 with a 1-sided alpha of 2.5%.

• Is an interim analysis necessary/efficient?
CO.20 Interim Analysis Simulations

General Assumptions

- Constant hazard rate(s) = exponential survival
- Linear rate of accrual and 1:1 randomisation
- 2 month lag of reporting of deaths on study + 3 month lag from trigger to interim analysis (data cleaning, analysis, DSMC report, etc.) = total lag of 5 months
- Target sample size = 750 patients
- Final analysis at 580 deaths**
- Median survival of cetuximab + placebo = 6 months
- Median survival of cetuximab + BMS-582664 = 6 months
- Proposed interim analysis at 290 events (1/2 of deaths required for final analysis)
CO.20 Interim Analysis Simulations

Scenario #1

- 30 patients per month accrued
CO.20 Interim Analysis Simulations

Scenario #3

- 50 patients per month accrued

Interim Analysis Results at 750 Patients Accrued, 267 Alive On Study
CO.20 Interim Analysis: Considerations

- “Mild” toxicity profile of Brivanib
- Availability of Cetuximab to non-trial patients (e.g. stopping early in Canada would potentially deprive any further patients from receiving cetuximab on study)
- Need to continue patients on cetuximab regardless of interim analysis results
- Small portion of final analysis alpha which must be "bought-out" to facilitate an interim analysis
A formal interim analysis for survival will be performed on all randomized subjects when at least 50% of the events (>263 deaths) have been observed, which is expected to occur approximately 17 months after the first patient is randomized. This analysis, based on the stratified logrank test adjusting for performance status (ECOG 0-1 vs. 2) at randomization, will test the following:

H₀: survival on brivanib (BMS-582664) + cetuximab ≤ survival on placebo + cetuximab

versus

H₁: survival on brivanib (BMS-582664) + cetuximab > survival on placebo + cetuximab

The comparison will be tested using the interim monitoring feature of EaSt software (Cytel Inc., Cambridge, MA, USA) based on a generalization of the Lan-DeMets error spending function approach using an O’Brien-Fleming stopping boundary to reject both H₀ and H₁, controlling for a one-sided alpha of 2.5% at the end of the study. For example, if exactly 263 deaths (50% of events) were in the locked database for the interim analysis, the nominal critical points for rejecting H₀ and H₁ would be respectively 2.767 and 0.438, corresponding to p-values of 0.0028 and 0.3308, respectively. Thus H₀ would be rejected early if the one-sided p-value from stratified log-rank test < 0.0028 and H₁ would be rejected early if the p-value > 0.3308.

Results of the interim analysis will be supplied to the DSMC who will communicate their recommendation regarding continuation of the trial to the Director of the NCIC CTG.
CONCLUSION

- Interim analysis plan should be carefully considered and prespecified in the protocol.
- DSMC infrastructure is important
  - terms of reference and reporting responsibility must be stated.
- DSMC represents patients interest on
  - accrual, consent, trial conduct, safety, efficacy, adequate evidence for changing practice.
Correlative Study Analyses

Examples
HER2 and Responsiveness of Breast Cancer to Adjuvant Chemotherapy

Kathleen I. Pritchard, M.D., Lois E. Shepherd, M.D., Frances P. O’Malley, M.D., Irene L. Andrulis, Ph.D., Dongsheng Tu, Ph.D., Vivien H. Bramwell, M.B., B.S., and Mark N. Levine, M.D., for the National Cancer Institute of Canada Clinical Trials Group
NCIC CTG-MA5
Pre-menopausal node positive (n=710)

CMF  6 cycles every 4 weeks
• Cyclophosphamide 100 mg/m² po x 14 d
• Methotrexate 40 mg/m² iv d 1 & 8
• 5FU 600 mg/m² iv d 1 & 8

CEF  6 cycles every 4 weeks
• Cyclophosphamide 75 mg/m² po x 14d
• Epirubicin 60 mg/m² iv d 1 & 8
• 5FU 500 mg/m² iv d 1 & 8
Cotrimoxazole or norfloxacin/ciprofloxacin
NCIC CTG MA. 5

- Patients accrued from 1989 to 1993
- First results published in 1998 which showed that CEF is superior to CMF in both relapse free and overall survivals
- FDA approved CEF for the treatment of early breast cancer in 1999
- CEF became a standard treatment in Canada for premenopausal women with node positive breast cancer
- CEF is however more toxic than CMF (associated with increased risk in heart failure and leukemia) and also more expensive
- There was a need for a biomarker which would be used to identify patients who will benefit from CEF
MA.5 Overall Survival

HR = 0.85  p = 0.047

CEF 351
CMF 359

Levine et al, JCO 2005
Correlative (translational) Studies in MA.5

- **HER2 overexpression by**
  - Immunohistochemistry with
    - CB 11 Antibody
    - TAB 250 Antibody

- **HER2 amplification by**
  - Polymerase chain reaction (PCR)
  - Fluorescence-in-situ hybridization (FISH)

- **All work carried out on paraffin embedded specimens**
Figure 1. Relapse-free Survival (Panel A) and Overall Survival (Panel B) among Women with Breast Cancer, According to HER2 Amplification Status on FISH.
Figure 2. Relapse-free Survival (Panel A) and Overall Survival (Panel B) According to the Type of Adjuvant Chemotherapy in Women with HER2 Amplification on FISH.
A

Relapse-free Survival (%)

No. at Risk

<table>
<thead>
<tr>
<th>Group</th>
<th>At Risk</th>
<th>Years 1</th>
<th>Years 2</th>
<th>Years 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEF</td>
<td>75</td>
<td>42</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>CMF</td>
<td>88</td>
<td>35</td>
<td>12</td>
<td></td>
</tr>
</tbody>
</table>

P = 0.003
Figure 3. Relapse-free Survival (Panel A) and Overall Survival (Panel B) According to Type of Adjuvant Chemotherapy in Women without HER2 Amplification on FISH.
A

Relapse-free Survival (%)

CEF

CMF

P = 0.49

Years

0 2 4 6 8 10

No. at Risk

CEF group 237 145 59
CMF group 228 138 60
### Adjusted* Hazard Ratios by HER2 Status (CEF vs. CMF)

<table>
<thead>
<tr>
<th>HER2</th>
<th>Relapse Free Survival</th>
<th>Overall Survival</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR</td>
<td>95% CI</td>
</tr>
<tr>
<td>Amplified</td>
<td>0.52</td>
<td>0.34 - 0.80</td>
</tr>
<tr>
<td>Not Amplified</td>
<td>0.91</td>
<td>0.71 - 1.18</td>
</tr>
</tbody>
</table>

* adjusted for age, nodal status, grade, ER status, surgical procedure, tumour size

Test for interaction: p=0.02 for DFS; p=0.01 for OS

Pritchard NEJ M 2006
Conclusions from MA.5 Correlative Analyses

• HER2 amplification or overexpression in breast cancer is associated with a larger benefit from CEF than CMF

• Patients whose tumours do not amplify or overexpress HER2 receive virtually no benefit from CEF, as compared to CMF

• Patients whose tumours do not exhibit HER2 amplification or overexpression could be treated with less toxic regimen of CMF

• Those with tumours which show amplified or overexpressed HER2 should receive dose-intense anthracycline-containing regiments such as CEF.
Limitations of MA. 5 Results to Clinical Practice
(From Editorial by Martine Piccart-Gebhart)

• A benefit of CEF to patients whose tumours do not amplify or overexpress HER2 cannot be firmly ruled out.

• It is now known from high-throughput gene-expression profiling of breast cancer that HER2 negative tumour includes at least three different subforms: basal-like; luminal B; luminal A.

• Chemotherapy may still be beneficial for HER2 negative patients with luminal B and basal-like breast cancer.
The Need for Better Biomarkers

• “It is thought provoking that after 30 years of modern tumour marker research, clinically useful cancer markers are still rare”

• “Gene expression profiling and other high-throughput genomic techniques are likely to find their own niche in the near future”

• Molecular signatures identified from genomics and proteomics studies could prove to be more “accurate” than a single gene biomarker since any particular gene that functions as part of a complex network may contain only limited information about the activity of the entire pathway.
A Multigene Assay to Predict Recurrence of Tamoxifen-Treated, Node-Negative Breast Cancer

Soonmyung Paik, M.D., Steven Shak, M.D., Gong Tang, Ph.D., Chungyeul Kim, M.D., Joffre Baker, Ph.D., Maureen Cronin, Ph.D., Frederick L. Baehner, M.D., Michael G. Walker, Ph.D., Drew Watson, Ph.D., Taesung Park, Ph.D., William Hiller, H.T., Edwin R. Fisher, M.D., D. Lawrence Wickerham, M.D., John Bryant, Ph.D., and Norman Wolmark, M.D.
Development of Oncotype DX™ 21-Gene Assay

- Development of a high-throughput, real-time, RT-PCR method to quantify gene expression with the use of sections of fixed, paraffin-embedded tumor tissue

- Selection of 250 candidate genes from published literature, genomic databases, and experiments based on DNA arrays performed on fresh-frozen tissue

- Analysis of data from three independent clinical trials of breast cancer to test the relationship between expression of the 250 candidate genes and the recurrence of breast cancer

- Selection of a panel of 16 cancer-related genes and 5 reference genes to generate an algorithm to calculate a recurrence score based on levels of expression of these genes
Gene Expression and Benefit of Chemotherapy in Women With Node-Negative, Estrogen Receptor–Positive Breast Cancer

Soonmyung Paik, Gong Tang, Steven Shak, Chungyeul Kim, Joffre Baker, Wanseop Kim, Maureen Cronin, Frederick L. Bachner, Drew Watson, John Bryant, Joseph P. Costantino, Charles E. Geyer Jr, D. Lawrence Wickerham, and Norman Wolmark
All Patients with RT-PCR Assay Data

A

Proportion Distant-Recurrent Free

Years

Tam + chemo
Tam

424  410 (7)  397 (10)  363 (18)  338 (24)  244 (30)  86 (32)
227  215 (6)  201 (17)  187 (22)  176 (24)  126 (26)  54 (30)

P = .02
Patients with Low Risk of Recurrence (Recurrence Score <18)

Hazard ratio: 1.31 (0.46, 3.78)
Patients with Intermediate Risk of Recurrence (Recurrence Score between 18 and 30)

Hazard ratio: 0.61 (0.24, 1.59)
Patients with High Risk of Recurrence (Recurrence Score higher than 30)

The p-value of the interaction test between RS and treatment = 0.038

Hazard ratio: 0.26 (0.13, 0.53)
Conclusions from RS and Chemotherapy Analysis

- Patients with tumours that had low recurrence score derived minimal, if any, benefit from chemotherapy treatment, while patients with tumours that had high recurrence score experienced a large chemotherapy benefit.

- Patients with tumours that had intermediate recurrence score did not appear to receive a substantial benefit, but the uncertainty in the estimate (relative risk=0.61 with 95% CI from 0.24 to 1.59) cannot exclude a clinically important benefit from chemotherapy treatment.

- The Oncotype DX 21 Gene Assay not only quantifies the likelihood of breast cancer recurrence in women with node-negative, estrogen receptor-positive breast cancer (i.e., as a prognostic marker), but also predicts the magnitude of chemotherapy benefit (i.e., as a predictive marker).
The Trial Assigning Individualized Options for Treatment (Rx), or TAILORx (N=10,046)

- **Secondary Study Group - 1**
  - Recurrence Score < 11 (~29% of Population)
  - Patients = Registered

- **Primary Study Group**
  - Recurrence Score 11-25 (~44% of Population)
  - Patients = Randomized

- **Secondary Study Group - 2**
  - Recurrence Score > 25 (~27% of Population)
  - Patients = Registered

**Stratify**
- Tumor Size: ≤ 2.0 cm vs. ≥ 2.1 cm
- Post-menopausal vs. Pre- or Peri-menopausal
- Planned chemotherapy: Taxane-containing (i.e. paclitaxel, docetaxel) vs. Non-taxane-containing
- Planned radiation therapy: whole breast, no boost planned vs. whole breast, boost planned vs. partial breast irradiation planned vs. no planned radiation therapy (for patients who have had a mastectomy)

- **Arm A**
  - Hormonal Therapy

- **Arm B**
  - Hormonal Therapy

- **Arm C**
  - Chemotherapy Plus Hormonal Therapy

- **Arm D**
  - Chemotherapy Plus Hormonal Therapy
Statistical Issues: Validation of Multivariate Index Predictive Markers

- What should be used to measure the accuracy of a predictive biomarker, especially with censored data?
- How to compare two different biomarkers developed from different sets of variables?
- We could use the coefficient and p-value for the interaction term in a Cox model but the proportional assumption may not be true.
- Nonparametric measurement of interactions?
- Randomized clinical trials are the best answer?
Phase III Randomized Study of 70-Gene Signature (Mammaprint™) Versus Clinical Assessment in Selecting Women With Node-Negative Breast Cancer for Adjuvant Chemotherapy

(MINDACT: Microarray In Node negative Disease may Avoid ChemoTherapy; N=6000)
Statistical Issues: Design of Studies for Multivariate Index Predictive Markers

- The sample size in a clinical trial, especially for earlier cancers, is usually very large but the collection of tissues and assays for the gene expressions may be very expensive.

- Can we use case-only, case-cohort, nested case-control, or other design so we don’t need to collect tissues and perform assays for all patients randomized?

- How much is the loss of efficiency if the primary objective is to identify a predictive marker?

- Best design of clinical trials to validate and compare predictive biomarkers?
Assessing Clinical Utility of a Predictive Marker

Clinical Trial Designs for Predictive Marker Validation in Cancer Treatment Trials

Daniel J. Sargent, Barbara A. Conley, Carmen Allegra, and Laurence Collette
**Indirect:** Marker by Treatment Interaction Design

Two independent clinical trials of Tx A *versus* Tx B

1. Separate tests of efficacy of treatment by group
   - Larger sample size required; Powered for efficacy assessment in each group
2. Formal statistical test of interaction
   - Smaller sample size required; Powered for single statistical test of interaction
Direct: Marker-Based Strategy Designs

- Standard treatment = Tx A
- Compare outcome of all marker-based *versus* all non-marker based patients
- Does not examine effect of Tx B (likely marker-based) in marker (-) patients = if Tx B is universally superior, regardless of marker status, this could not be determined
Direct: Marker-Based Strategy Designs

- Second randomization allows clarification of whether any effect is due to true effect of marker status, or superiority of Tx B regardless of marker status.
- Direct designs may be preferred for:
  - Multiple/panel of markers
  - Multiple treatments
  - Multiple efficacy outcomes
• “Investigation of predictive effects for a marker is, by definition, a prospective subset analysis: in other words, does the treatment effect differ in subgroups defined by a marker level. Therefore, a larger sample size is necessary”