Development of omics-based tests for clinical use: the challenge of achieving statistical robustness and clinical utility

*FDA Proteomics in the Clinic Workshop*

Lisa McShane, PhD
Biometric Research Branch, DCTD
National Cancer Institute

June 13, 2014
Disclosures

• I have no financial relationships to disclose.
• I will not discuss off label use and/or investigational use in my presentation.
• The views expressed represent my own and do not necessarily represent views or policies of the National Cancer Institute.

My perspective

• Statistical/scientific reviewer of NCI-sponsored studies for development and validation of biomarker-based tests
• Scientific Advisory Board (Science Translational Medicine) and Editorial Board (BMC Medicine)
• Statistical collaborator in research projects
OUTLINE

• Background & definitions
• Roles for omics-based tests
• Define prognostic and predictive
• Two cases studies
  • Gene expression-based prognostic classifier in early stage lung cancer
  • Serum proteomic predictive classifier in advanced lung cancer
• Recommended reading
Working definitions

• Biomarker
  (http://www.cancer.gov/dictionary):
  “Biological molecule found in blood, other body fluids, or tissues that is a sign of a normal or abnormal process, or of a condition or disease.”

• Omics
  (http://www.iom.edu/Reports/2012/Evolution-of-Translational-Omics.aspx)
  “A term encompassing multiple molecular disciplines, which involve the characterization of global sets of biological molecules such as DNAs, RNAs, proteins, and metabolites.”
Many examples of biomarkers/omics for characterization of biological samples

- **Illumina SNP bead array**
- **Affymetrix expression GeneChip**
- **MALDI-TOF proteomic spectrum**
- **cDNA expression microarray**
- **Mutation sequence surveyor trace**
- **p53 IHC stain of breast cancer**
- **FISH analysis of BCR-ABL in ALL**
- **SKY analysis of AML cells**
Potential roles for omics/biomarker-based tests

- **Pre-diagnosis**
  - Risk
  - Screening
  - Early detection

- **Pre-treatment**
  - Prognostic*
  - Predictive (treatment-selection)*

- **Intra-treatment**
  - Early response or futility
  - Toxicity monitoring

- **Post-treatment**
  - Early endpoint
  - Recurrence or progression monitoring

*Examples in this talk focus on tests for initial therapy selection.
Paradigm for development of a clinically useful biomarker-based test

Discovery

Clinical validity
The test result shows an association with a clinical outcome of interest.

Analytical validity
The test’s performance is established to be accurate, reliable, and reproducible.

Clinical utility
Use of the test results in a favorable benefit to risk ratio for the patient.

Genet Med 2009;11:3-14
J Clin Oncol 2012;30:4223-4232
Prognostic biomarker

- Associated with clinical outcome in absence of therapy (natural course) or with standard therapy all patients are likely to receive
- Not always relevant for therapy decisions

Good prognosis group (M-) may forego additional therapy

Is this prognostic information helpful?

Hazard ratio = .18

Hazard ratio = .56
Predictive biomarker

• Associated with benefit or lack of benefit (potentially even harm) from a particular therapy relative to other available therapy
  • Alternate terms: treatment-selection, treatment-guiding, treatment effect modifier
• Generally more useful than prognostic biomarkers for therapeutic decision making

Clinical Trials 2013; 10: 653-665
Prognostic vs. predictive: Importance of control groups

- Prognostic but not predictive
- (M = biomarker)
- Prognostic and predictive
Statistical language for predictive biomarkers: “Treatment-by-biomarker interaction”

- Treatment effect (e.g., hazard ratio) varies by biomarker status
  - **Quantitative** interaction: Treatment benefits all patients but by different amounts
  - **Qualitative** interaction: Patients “positive” for the biomarker benefit from the treatment but others receive no benefit or possibly even harm
Plasma IL-6 as predictive biomarker for pazopanib vs. placebo?

Results of randomized placebo-controlled phase III trial in metastatic renal-cell cancer

<table>
<thead>
<tr>
<th>Interleukin 6</th>
<th>PFS (weeks)</th>
<th>HR (95% CI)</th>
<th>p value</th>
<th>Predictive?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>42.3</td>
<td>0.55 (0.38–0.81)</td>
<td>0.445</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>High</td>
<td>32.6</td>
<td>0.31 (0.21–0.44)</td>
<td>0.009</td>
<td></td>
</tr>
</tbody>
</table>

Prognostic: P<0.0001  
Quantitative interaction: P=0.009

EGFR mutation predictive for PFS benefit with gefitinib in NSCLC

EGFR mutation is:
- Prognostic (positive)
- Predictive: Qualitative interaction, p<0.001

**IPASS:** Phase III 1\textsuperscript{st} line advanced adeno NSCLC gefitinib vs. carboplatin+paclitaxel (N Engl J Med 2009;361:947-57)

Cessation of chemo?

\[
\text{HR}=0.74 \quad P<0.001
\]

\[
\text{HR}=0.48 \quad P<0.001
\]

\[
\text{HR}=2.85 \quad P<0.001
\]
A 15-gene signature was constructed using data from OBS arm of a randomized clinical trial (OBS vs. ACT) for lung cancer patients who were candidates for adjuvant chemotherapy.

“A 15-gene signature separated OBS patients into high-risk and low-risk subgroups with significantly different survival (hazard ratio [HR], 15.02; 95% CI, 5.12 to 44.04; \( P < .001 \)).”

(J Clin Oncol 2010; 28: 4417-4424)
A statistical model is **OVERFIT** when it describes random error or noise instead of the true underlying relationship.

- Excessively complex (too many parameters or predictor variables)
- Generally has poor predictive performance on an independent data set

**RESUBSTITUTION** is the naïve practice of evaluating performance of a model by "plugging in" exact same data used to build it.
Model development

Model “resubstitution” pitfall


- Goal: Develop prognostic signature from gene expression microarray data
- Survival data on 129 lung cancer patients (prior study)
- Expression values for 5000 genes generated randomly from $\mathcal{N}(0, 1_{5000})$ (“noise”) for each patient
- Data divided randomly into training and validation sets
- Prognostic model developed from training set and used to classify patients in both training and validation sets (supervised principal components method)
Prognostic classifier for early stage non-small cell lung cancer

Did it really validate?

“. . . prognostic effect was validated consistently in four separate microarray data sets (total 356 stage IB to II patients without adjuvant treatment).”

- What happened to HR=15.02?
- Endpoint: DSS→OS
- Timescale: 9 →5 yrs
- Mixed stages
The signature was also predictive of improved survival after ACT in JBR.10 high-risk patients (HR, 0.33; 95% CI, 0.17 to 0.63; P = .0005), but not in low-risk patients (HR, 3.67; 95% CI, 1.22 to 11.06; P = .0133; interaction P < .001).” (J Clin Oncol 2010; 28: 4417-4424)

RESUBSTITUTION strikes again
Model development: Serum proteomic test to classify NSCLC for outcome with EGFR-TKIs

- Serum collected from NSCLC patients before treatment with gefitinib or erlotinib (EGFR-TKIs)
- Analysis by MALDI-MS
- K-nearest neighbor (KNN) algorithm based on 8 distinct m/z features classifies into good or poor outcome
- Training set: n=139 NSCLC patients total from 3 cohorts who received gefitinib
- Preliminary validation cohorts:
  - “Italian B”: n=67 sequential patients, late-stage or recurrent NSCLC treated with single-agent gefitinib
  - ECOG 3503: n=96 advanced NSCLC patients treated with first-line erlotinib on single arm Phase II study

Preliminary validation: Proteomic test to classify NSCLC for outcome with EGFR-TKIs

Preliminary results for patients treated with EGFR-TKIs

"Italian B": n=67 sequential patients, late-stage or recurrent NSCLC treated with single-agent gefitinib
HR=0.50, 95% CI=(0.24,0.78), p=0.0054
Median OS
Good: 207 days  Poor: 92 days

ECOG 3503: n=96 advanced NSCLC patients treated with first-line erlotinib on single arm Phase II study
HR=0.4, 95% CI=(0.24,0.70), p<0.001
Median OS
Good: 306 days  Poor: 107 days

Proteomic test shown to have good analytical reproducibility across 2 labs
Predictive or Prognostic? Proteomic test to classify NSCLC for outcome with EGFR-TKIs

Does test also separate by outcome patients who did NOT receive EGFR-TKIs (control cohorts)?

“Italian C”: n=32 patients, stage IIIA-IV NSCLC treated with second-line chemotherapy
HR=0.74, 95% CI=(0.33,1.6), p=0.42

SAME TREND, BUT NS

“VU”: n=61 patients, advanced NSCLC treated with second-line chemotherapy
HR=0.81, 95% CI=(0.4,1.6), p=0.54

SAME TREND, BUT NS

“Polish”: n=65 patients, stage IA-IIB NSCLC treated with second-line chemotherapy
HR=0.90, 95% CI=(0.43,1.89), p=0.79

SAME TREND, BUT NS
Randomized phase III trial (PROSE): Proteomic test to classify NSCLC for outcome with EGFR-TKIs

- Test predictive value of the proteomic test
- Primary endpoint overall survival (OS)
- Powered for treatment x proteomic test interaction

Eligibility
- Stage IIIIB or IV NSCLC
- ≥ 18 years old
- Refractory to one prevision platinum-containing regimen

Exclusions
- Previously received an EGFR-TKI
- Uncontrolled brain metastases
- Other cardiac, renal, etc. conditions
Randomized phase III trial (PROSE): Proteomic test to classify NSCLC for outcome with EGFR-TKIs

"Serum protein test status is predictive of differential benefit in overall survival for erlotinib versus chemotherapy in the second-line setting. Patients classified as likely to have a poor outcome have better outcomes on chemotherapy than on erlotinib.” (Lancet Oncol 2014;15:713-21)
Randomized phase III trial (PROSE): Proteomic test to classify NSCLC for outcome with EGFR-TKIs

The indication for the test seems to have drifted from a test to select who will benefit from erlotinib to who should receive chemotherapy.
Proteomic test to classify NSCLC for outcome with EGFR-TKIs: Many questions remain

- Impact of patient selection criteria for trial (patients could not have prior EGFR-TKI)
- Impact of subsequent therapies on OS endpoint
- Important differences in drug delivery (oral vs. IV)
- Important differences in toxicity profile
- Is giving all patients chemotherapy a reasonable option?
Institute of Medicine report on the field of translational omics

NCI criteria for the use of omics-based predictors in clinical trials.

Thanks for your attention!