Statistical Challenges in Precision Medicine with a Focus on Companion Diagnostics

Gregory Campbell, Ph.D.
President, GCStat Consulting
GCStat@verizon.net
Former Director, Division of Biostatistics,
Center for Devices and Radiological Health, FDA

Biomarker Symposium, ASA P-T, Rutgers, NJ
Outline

1. Types of Genetic Tests and Biomarkers
2. Regulation of *In Vitro* Diagnostics
3. Statistical Evaluation of *In Vitro* Diagnostics
4. *In Vitro* Companion Diagnostics
5. Co-development of a Predictive Diagnostic and a Therapeutic
6. Bridging Studies
7. Complementary Diagnostic Tests
8. Some Challenges
Terminology

• Personalized Medicine
• Stratified Medicine
• Tailored Therapeutics
• Precision Medicine

• “the right medical product to the right patient at the right time”
1. Some Biomarker Technologies

- Immunohistochemistry (IHC)
- Fluorescence in Situ hybridization (FISH)
- Genomics
  - A test based on one Single Nucleotide Polymorphism (SNP) or Polymerase Chain Reaction (PCR)
  - Profile of a number of SNPs
  - Gene expression at a single location
  - Microarrays for gene expression
  - Next Generation Sequencing (NGS)—high through-put
- Protein Expression
Types of Genetic Tests

● Single Nucleotide Polymorphisms (SNPs)
  ● Basically Qualitative Assay: Is the particular sequence present or not. Examples include Factor V Leiden, HLA typing, cytochrome P-450 superfamily SNPs Microarrays
  ● Basically quantitative, measuring gene expression

● Next Generation Sequencing (NGS)
  ● Detect DNA changes in CFTR (cleared Nov. 2013)

Next Generation Sequencing (NGS)

• In 2001 the sequencing of the 3 billion base pairs of a human genome was announced. It took more than 10 years and cost $3 billion.

• NGS is massively parallel, high through-put

• In 2014 Illumina announced its NGS machine Hi Seq X Ten can sequence 45 human genomes in one day at a cost of $1000 each.

• In 2013 FDA cleared (authorized the marketing of) one NGS instrument (Illumina MiSeqDx) and its universal sequencing reagents, and two accompanying assays for the diagnosis of cystic fibrosis.
2. Federal Regulation of In Vitro Diagnostic (IVD) Tests

- 1967 Clinical Laboratory Improvement Act (CLIA)
- 1972 Proposal for FDA to regulate microbial sensitivity diagnostics as drugs
  - Uniform labeling proposed instead of requiring NDA
- 1976 Medical Device Amendments to the Food, Drug and Cosmetic Act]
  - Pre-market Notification -- 510(k)
  - Pre-market Approval -- PMA
- 1988 CLIA Amendments expand the scope beyond high complexity labs (6000) to all labs including physicians’ offices (150,000)
CLIA vs. FDA Review

• A 510(k) review is based on assessment of “substantial equivalence” to another device and may include performance data.

• A CLIA review is based on complexity of testing with no review of performance data.

• Complexity categorization is where a new cleared 510(k) IVD is assigned:
  • moderate or high

• Device can be waived by CLIA if assessed as simple (in hands of lay user).
## In Vitro Diagnostic (IVD) Settings

<table>
<thead>
<tr>
<th>Setting</th>
<th>IV Diagnostic</th>
<th>Regulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>In-house assay</td>
<td>In house assay</td>
<td>None</td>
</tr>
<tr>
<td>Single Site IV Clin. Diagnostic</td>
<td>Single site assay</td>
<td>FDA CLIA</td>
</tr>
<tr>
<td>Commercial Reagents</td>
<td>Analyte Specific Reagents</td>
<td>FDA (510(k) or de novo (dn))</td>
</tr>
<tr>
<td>High Complexity Lab Kits</td>
<td>Regulated IVDs</td>
<td>FDA (510(k), dn, or PMA) CLIA High Compl. Labs</td>
</tr>
<tr>
<td>Point-of-Care</td>
<td>CLIA Waived tests</td>
<td>FDA (510(k), dn, or PMA)</td>
</tr>
<tr>
<td>Self-Care</td>
<td>OTC IVD’s</td>
<td>FDA (510(k), dn, or PMA)</td>
</tr>
</tbody>
</table>
IVD Diagnostic Tests Used in Cancer

• Require a PMA
3. Diagnostic Devices

- Can be used for
  - Diagnosis
  - Screening
  - Monitoring disease or medical condition

- Types of devices
  - *In vitro* diagnostic devices
  - Imaging systems
  - Other *in vivo* devices
Diagnostic Test Evaluation

• Preclinical (analytical) evaluation
  • Repeatability, Precision (coefficient of variation and variance components)
  • Limit of Detection (LOD) of the biomarker
  • Linearity for dilution recovery
  • Traceability
  • Interference
  • Stability
  • Cross reactivity

• Clinical performance evaluation
  • Method comparison
  • Clinical sensitivity and specificity, if clinical reference standard exists, and agreement to another approved test (comparator) if not.
Statistical Methodology for Clinical Evaluation of a Diagnostic Test

• If there is a Clinical Reference Standard (CRS)
  (A clinical reference standard is “considered to be the best available method for establishing the presence or absence of the target condition”)
  • Sensitivity and Specificity
  • Positive and Negative Predictive Values
  • Receiver Operating Characteristic (ROC) Methodology

• If no CRS, the focus is on agreement of a new test with a comparator:
  • Method Comparison
  • Positive and Negative Agreement
Diagnostic: Analyses Using ROC Plots

• Since the Receiver Operating Characteristic (ROC) plot is a plot over all sensitivities and specificities, it gives a global assessment, a visual presentation of the entire performance
• Very useful methodology in CDRH
• If the data are ordinal, one can use latent variables to build the theoretical ROC curve
• ROC methodology can be used in a variance components effort to model the variance due to readers, to cases and help plan the trial.
CLSI

• International Standards Organization that develops global consensus standards and guidelines for health care testing with representatives from industry, government, and professional organizations

• CLSI Evaluation Protocols (EP) Committee focuses on study design and analysis (statistics)
Relevant CLSI Evaluation Protocol (EP) Documents

- EP5  Evaluation of Precision Performance of Quantitative Measurement Methods
- EP9  Method Comparison and Bias Estimation Using Patient Samples
- EP12  User Protocol for Evaluation of Qualitative Test Performance
- EP17  Limits of Detection and Limits of Quantitation
- EP21  Total Analytical Error
- EP24  Assessment of the Clinical Accuracy of Laboratory Tests Using Receiver Operating Characteristic (ROC) Plots
Guidelines for Reporting Diagnostic Test Results

• STARD Initiative for reporting studies of diagnostic accuracy in medical journals
    Also in Ann. Int. Med. and others


http://www.fda.gov/MedicalDevices/DeviceRegulation andGuidance/GuidanceDocuments/ucm071148.htm
4. FDA Guidance

Issued jointly by CDRH, CDER and CBER on August 6, 2014.

IVD Companion Diagnostic Device

• An *in vitro* diagnostic device that provides information that is **essential** for the safe and effective use of a corresponding therapeutic product (could be a drug, biologic or another device).

• Its use is stipulated in the instructions for use in the labeling of both the diagnostic device and the therapeutic product. The therapeutic label may say “as determined by an FDA-approved test” and the label for the test will refer back to the therapeutic product.)
IVD Companion Diagnostic Device

• It could be used to:
  • Identify patients who are most likely to benefit from the product
  • Identify patients who are likely to be at increased risk for serious adverse reactions from the product
  • Monitor the response to treatment for the purposes of adjusting treatment
  • Identify patients in the population for whom the therapeutic product has been adequately studied, and found safe and effective, i.e., there is insufficient information about the safety and effectiveness of the therapeutic product in any other population.
Types of Biomarkers

1. Diagnostic—identify a disease or condition
2. Early Detection—identify a disease in an asymptomatic population
3. Monitoring patients with a disease
4. Risk assessment
5. Prognostic—better or worse outcome under standard therapy
6. Predictive for safety/Drug dosing
7. Predictive for Effectiveness (for a particular therapy)

These last three (5, 6, 7) concern prognosis and prediction and attempt to predict the future and as such rely on statistics.

Companion diagnostics focus on the last two (6, 7) which attempt to predict the future for therapeutic drugs.
Prognostic vs. Predictive

- (Also see Sargent et al (2005) J Clin Oncol 23: 2020-2027.)
Safety Biomarkers

• Predict risk of an adverse event dependent on the biomarker

• Example: Test to minimize neutropenia for cancer patients undergoing chemotherapy for colorectal cancer with irinotecan
  • UGT1A1 is a polymorphic enzyme and variant 28 of the allele has been found to predict heightened risk of neutropenia based on an understanding of the pathway for the drug in UGT1A1s
  • 30% of Caucasians have this variant
  • Invader UGT1A1 Molecular Assay by Third Wave Technologies was cleared by FDA on August 22, 2005.
Safety Biomarker for Dosing

• Warfarin is a blood thinner (to prevent clotting) but can lead to internal bleeding if not dosed correctly. Variants of a cytokine P-450 CYP2C9 and VKORC1 can help to decide on the correct dosing strategy.

• Nanosphere Verigene Warfarin Metabolism Nucleic Acid Test cleared by FDA Sept. 17, 2007

• It provides a different dosing strategy for those who test positive.
5. Predictive Biomarker

Marker: Her2-neu FISH (Hercept Test)
Treatment: trastuzumab (Herceptin)
Objective response rates:

<table>
<thead>
<tr>
<th></th>
<th>Herceptin+</th>
<th>Chemo Alone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemo</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FISH+</td>
<td>95/176 (54%)</td>
<td>51/168 (30%)</td>
</tr>
<tr>
<td>FISH-</td>
<td>19/50 (38%)</td>
<td>22/57 (39%)</td>
</tr>
</tbody>
</table>

Response rate advantage of trastuzumab (Herceptin) depends on Her2-neu FISH test outcome

Arch. Pathol. Lab Med Jan 2007 (ASCO/CAP Guidelines)
Predictive Biomarker

• Her2-neu (c-erb-B2) is predictive for trastuzumab (Herceptin).
• It is also prognostic for breast cancer recurrence.
• There are now at least 10 commercial tests for the oncogene Her2-neu (and they do not all completely agree) to guide trastuzumab (Herceptin)
Not All Her2-neu Tests Agree

Table 3  Comparison of 3 Her2-neu tests: IHC versus two FISH (INFORM™ and PathVysion™) assays

<table>
<thead>
<tr>
<th>FISH</th>
<th>Subtotal</th>
</tr>
</thead>
<tbody>
<tr>
<td>INFORM</td>
<td>Amplified</td>
</tr>
<tr>
<td>PathVys.</td>
<td>Amplified</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IHC</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>High +</td>
<td>10 (19%)</td>
<td>0</td>
<td>0</td>
<td>3 (6%)</td>
<td>13 (25%)</td>
</tr>
<tr>
<td>Med +</td>
<td>3 (6%)</td>
<td>1 (2%)</td>
<td>0</td>
<td>9 (17%)</td>
<td>13 (25%)</td>
</tr>
<tr>
<td>Low +</td>
<td>1 (2%)</td>
<td>0</td>
<td>0</td>
<td>8 (15%)</td>
<td>9 (17%)</td>
</tr>
<tr>
<td>Neg</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>17 (33%)</td>
<td>17 (33%)</td>
</tr>
</tbody>
</table>

**Subtotal** | 14 (27%) | 1 (2%) | 0 | 37 (71%) | 52 (100%) |

21 of 52 disagreed on at least one of the three tests, 2 FISH (fluorescence in situ hybridization) and one IHC (immunohistochemistry).

Predictive Markers for Efficacy

• Predictive biomarker predicts differential effect of a particular treatment on an outcome (such as survival, response, or recurrence)

• Statistically, there is an interaction between the biomarker and the particular treatment.

• Always need to include the phrase “predictive for a particular (drug) therapy”.

• Need to show that it works for test positive patients better than for test negative ones
Three Simple Trial Designs for a Efficacy of a Drug and a Diagnostic Test

• All Comers
  • This may be a very large trial if the prevalence of the test is small
  • The diagnostic test could be done retrospectively

• Biomarker-Stratified Design
  • Differentially oversample the rarer group (usually test positive); allows for assessment of treatment by marker interaction

• Enrichment (Targeted) Design
  • Study only the test positive patients
  • Can only obtain PPV and not sensitivity, specificity nor NPV
Types of Companion Diagnostic Devices

• A novel device
• A new version of an existing device by a different manufacturer
• An existing device that has been approved or cleared for another indication
• A new version of an existing device by the same manufacturer
Timing of Therapeutic Product and a Companion Diagnostic

A. New IVD companion diagnostic is usually developed contemporaneously and approved simultaneously with a novel therapeutic product

B. Existing diagnostic test for a new drug

C. New diagnostic test for an approved drug
A. Simultaneous Approval of New Drug and New Diagnostic

- Vysis ALK Break Apart FISH Probe Kit by Abbott Molecular (PMA)

- Intended Use: The Vysis ALK Break Apart FISH Probe Kit is a qualitative test to detect rearrangements involving the ALK gene via fluorescence in situ hybridization (FISH) in formalin-fixed, paraffin-embedded (FFPE) non-small cell lung cancer (NSCLC) tissue specimens to aid in identifying patients eligible for treatment with crizotinib. This is for prescription use only.

Approved 8/26/11
B. Existing Diagnostic Test for New Drug Example

• PMA for new Dako EGFR test approved contemporaneously with the new drug cetuximab in 2004.

• Dako EGFR PharmDx Kit approved in 2006 in PMA Supplement also for the new drug panitumab in 2006:
  • The EGFR pharmDx™ assay is a qualitative immunohistochemical (IHC) kit system to identify epidermal growth factor receptor (EGFR) expression in normal and neoplastic tissues routinely-fixed for histological evaluation EGFR pharmDx specifically detects the EGFR (HER1) protein in EGFR-expressing cells.
  • EGFR pharmDx is indicated as an aid in identifying colorectal cancer patients eligible for treatment with cetuximab or panitumumab.
Existing IVD Diagnostic Test

• If IVD diagnostic test is already legally marketed, but the manufacturer intends to market it for its new use with the therapeutic product then, as a new intended use, a new PMA or 510(k) would be required.
C. New Diagnostic Test for an Approved Drug

*therascreen* KRAS RGQ PCR Kit

**Intended Use:**

The *therascreen* KRAS RGQ PCR Kit is a real-time qualitative PCR assay used on the Rotor-Gene Q MDx instrument for the detection of seven somatic mutations in the human KRAS oncogene, using DNA extracted from formalin-fixed paraffin-embedded (FFPE), colorectal cancer (CRC) tissue. The *therascreen* KRAS RGQ PCR Kit is intended to aid in the identification of CRC patients for treatment with Erbitux® (cetuximab) based on a KRAS no mutation detected test result.

PMA P110030 approved July 6, 2012

It resulted in a labeling change for cetuximab
C. Kaplan Meier for OS for Wild Type

![Kaplan Meier Curve for OS](image)

**SUBJECTS AT RISK**
- Cet+BSC: 117
- BSC: 128

**PROPORTION ALIVE**

**STRATIFIED LOGRANK P-VALUE = 0.0017**

**MONTHS**
- 0, 1, 2, ..., 24

**CENSORED**
- Cetuximab + BSC
- BSC
C. Kaplan Meier for OS for Mutation Positive

![Graph showing Kaplan Meier curves for OS for Mutation Positive]
Co-Development Draft Guidance

• Issued July 15, 2016
• Jointly issued by CDRH, CDER and CBER
• Discuss strategies and designs
Two Pivotal Clinical Study Designs for Companion Diagnostics

• If the test is commercially ready, pivotal clinical study to validate both the therapeutic product and the diagnostic. (This is ideal.) Or the diagnostic test is applied retrospectively in a prospective manner.

• Otherwise, therapeutic product is validated in a separate clinical trial from the diagnostic test. In this case, usually the therapeutic trial is first. Then a bridging study may be required since a clinical trial assay (CTA) was used in the clinical trial rather than a market ready test (MRT).
6. IVD Bridging Study

• Not bridging studies in Multi-regional Clin. Trials.
• A market ready test (MRT) for companion diagnostic test (CDx) may not be available at the time of the pivotal clinical trial of the drug and so a (non-commercial) clinical trial assay (CTA) is used instead.
• So a bridging study is a supplemental agreement study of MRT and CTA. It is designed to assess the agreement between MRT and CTA and can allow for the extrapolation of CTA’s clinical data to MRT. How much should they agree?
IVD Bridging Studies Challenges

• If there is genetic material from the trial:
  • It still may be difficult to re-consent the subjects for the new test MRT. If only some can be tested with MRT, this becomes a convenience sample (missing data!) and could introduce a bias of unknown size.
  • The specimen quality may have deteriorated due to handling or storage or the specimen material may be inadequate.
  • If RCT is only on CTA+ subjects, then no subjects who are CTA- but MRT+ can be studied, creating a bias of unknown size with an inability to estimate the therapeutic effect in the MRT+ subjects.

• If no genetic material is available, a separate agreement study between CTA and MRT may be performed.
IVD Bridging Studies Statistical Analysis Plan

• Evaluate discordance
• Treatment of missing samples
• Address (statistical) bias
• Consider the impact of efficacy of the drug using the MRT (hypothesis test (if appropriate) and estimate)
Companion Diagnostics

• As of Jan. 2017, there have been 30 Companion Diagnostic (CDx) tests approved or cleared by FDA and 2 Complementary Diagnostic tests. For the complete list of each companion diagnostic test, drug and indication for the combined use, http://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/InVitroDiagnostics/ucm301431.htm

• Of CDx tests, IHC (10), PCR (10), ISH (8), NGS (1), MRI (1). 16 used CTA and 14 used bridging studies. All but two are for oncology drugs.
7. Complementary Diagnostic Test

• A complementary diagnostic test is a test that aids in the benefit-risk decision-making about the use of the therapeutic product, where the difference in benefit-risk is clinically meaningful. Complementary IVD information is included in the therapeutic product labeling.

• Example: PD-L1 IHC 28-8, complementary for nivolumab for patients with melanoma and NSCLC

• Example: PD-L1 (SP-142), an IHC, for bladder cancer and NSCLC
Nivolumab vs. Docetaxel

• Randomized, double blind, Phase 3 study comparing nivolumab (a PD-L1 inhibitor) versus docetaxel for patients with advanced or metastatic NSCLC.


• PD-L1 IHC 28-8 pharmDx by Dako approved by FDA as complementary diagnostic for nivolumab for NSCLC patients on Oct. 8, 2015 (also approved for melanoma on Oct. 9, 2015).
Overall Survival, Patients >1% PD-L1
Overall Survival, Patients <1% PD-L1
Forest Plot OS Based on PD-L1

<table>
<thead>
<tr>
<th>PD-L1 expression level</th>
<th>Unstratified HR</th>
<th>Median OS (months) OPDIVO</th>
<th>Median OS (months) Docetaxel</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥1% (n = 246)</td>
<td>0.59</td>
<td>17.1</td>
<td>9.0</td>
</tr>
<tr>
<td>&lt;1% (n = 209)</td>
<td>0.90</td>
<td>10.4</td>
<td>10.1</td>
</tr>
<tr>
<td>≥5% (n = 181)</td>
<td>0.43</td>
<td>18.2</td>
<td>8.1</td>
</tr>
<tr>
<td>&lt;5% (n = 274)</td>
<td>1.01</td>
<td>9.7</td>
<td>10.1</td>
</tr>
<tr>
<td>≥10% (n = 165)</td>
<td>0.40</td>
<td>19.4</td>
<td>8.0</td>
</tr>
<tr>
<td>&lt;10% (n = 290)</td>
<td>1.00</td>
<td>9.9</td>
<td>10.3</td>
</tr>
</tbody>
</table>
Companion or Complementary Diagnostic Approach?

• Complementary diagnostic approach necessitates all comers or stratified biomarker design (not enrichment)
• It does not seem necessary to pre-specify a cutoff but to allow the data to inform the use of the diagnostic.
• The drug should provide benefit-risk advantage even for all patients.
8. Statistical Challenges for IVDs

• Precision studies (repeatability and reproducibility)
• Multiplicity of biomarkers and classifiers, and of models
• Subgroup identification
• Selection of threshold (cutoff)
• Training set and independent validation set
• Importance of independent validation of locked-down biomarker and threshold (Dangers of cross-validation or even random data splits)
• Use of retrospective data in a very limited and prospective manner (“prospective-retrospective”)
• Missing data (“intention to diagnose”) (see Campbell, Pennello & Yue (JBS, 2011))
More Statistical Analysis Challenges

• Analysis challenges
  • Subgroup analyses “If you torture the data long enough it will tell you anything you want”
  • Multiplicity
  • Use of retrospective data (prospectively)
  • Biomarkers as surrogates—build predictive models
Biological Challenges

• It is surprising how little we know about the biology.
• It is usually never as simple as you think it is.
• Examples
  • Many polymorphisms for cytochromes P-450
  • Variability of Her2-neu tests (and others)
  • Only certain alleles (1, 28) of ATG1A1 gene
  • EGFR and then kras
  • 7 somatic mutations for kras
Discuss with FDA

• Come in a meet with therapeutic and IVD review divisions (CDER and CDRH) before launching any registration trials for tailored therapeutics and take advantage of pre-submission meetings for CDRH and pre-NDA/pre-BLA meetings with CDER/CBER.
The Great Divide

• Many pharmaceutical drug companies do not know diagnostics and diagnostic companies are not familiar with therapeutic clinical trials methodology.

• Most biostatistics departments teach clinical trials but very few teach the statistical evaluation of diagnostics tests.

• The number of statisticians who understand both worlds is extremely small.

• In Division of Biostatistics in CDRH, two of the five branches are diagnostic, one \textit{in vitro} and imaging and the other \textit{in vivo}. 
Genomics and Precision Medicine

• Even if the therapeutic drug or device is precise, the diagnostic medical test that identifies what subgroup a patient belongs may not be.

• There can be large variability in the performance of diagnostic tests for the same condition (or disease). For example, tests for Her2neu do not all give the same results on the same set of patients.

• A question every statistician should ask is how variable and how accurate are the results of such diagnostic tests? There is a need for precision diagnostics.
Summary

• Personalized medicine is already here and progress will accelerate in the future as we come to a better understanding of the biology, develop more powerful informatic tools and follow good statistical principles for design and analysis.
Summary

• There are a great many challenges to the development of diagnostic tests

• Optimism that, as we gain more understanding of the biology, better diagnostic tests will be developed.

• Tailored (“personalized”) medicine is already here and progress will accelerate in the future as we come to a better understanding of the biology, develop more powerful informatics tools and follow good statistical principles for design and analysis of clinical studies.

• Meet with both CDER and CDRH early.
Questions?