Statistical Methods for Evaluating Diagnostic Devices

Changhong Song, FDA/CDRH

Disclaimer:

This presentation reflects the views of the author and should not be construed to represent FDA's views or policies.

Outline

- CDRH and DBS
- Introduction to diagnostic devices
- Study design and statistical analysis methods for evaluating diagnostic devices
- Application of Bayesian methods in medical device clinical studies

FDA/CDRH

- The Center for Devices and Radiological Health (CDRH) is the branch of the United States Food and Drug Administration (FDA) responsible for the premarket approval of all medical devices, as well as overseeing the manufacturing, performance and safety of these devices.

Division of Biostatistics

- Division of Biostatistics (DBS) is in the Office of Surveillance and Biometrics (OSB) in CDRH.
- Directors: Ram Tiwari; Lilly Yue; Yunling Xu
- DBS has 5 branches.
  - Therapeutic Statistics Branch 1
  - Therapeutic Statistics Branch 2
  - Therapeutic Statistics Branch 3
  - Diagnostic Statistics Branch 1
  - Diagnostic Statistics Branch 2 (I am here)
- Diagnostic Statistics Branch 2 Covers Office of In Vitro Diagnostics and Radiological Health (OIR), including IVDs and Dx Imaging devices)

Medical Device Evaluation

- There is reasonable assurance that a device is
  - **safe** when it can be determined, based upon valid scientific evidence, that the probable benefits to health from use of the device for its intended uses and conditions of use …outweigh any probable risks. 21 CFR 860.7(d)(1)
  - **effective** when it can be determined, based upon valid scientific evidence, that in a significant portion of the target population, the use of the device for its intended uses and conditions of use, … will provide clinically significant results. 21 CFR 860.7(e)(1)

Risk-Based Regulation

- **Class I:** low risk, simple
  - *Most exempt from pre-market review.*
  - *General controls are sufficient.*
- **Class II:** moderate risk, more complex
  - *Pre-market notification or "510(k)".*
  - *Special controls (per FDA guidance)*
- **Class III:** high risk, most complex
  - *Pre-market approval (PMA).*
  - *Safety, effectiveness*

Diagnostic Devices

- Diagnostic devices are described broadly as devices that provide results that are used alone or with other information to help assess a subject's health condition of interest, or target condition.
- Diagnostic devices include in in-vivo and In vitro diagnostic (IVD) devices.

Intended Uses for Diagnostic Devices

- **Diagnosis,** in symptomatic patients.
- **Screening**, in asymptomatic patients.
- **Early detection**, enabling intervention at an earlier and potentially more curable stage than under usual clinical diagnostic conditions.
- **Monitoring**, e.g., of disease response during therapy, with potential for adjusting level of intervention (e.g. dose) on a dynamic and personal basis.
- **Prognosis**, allowing for more (less) aggressive therapy for patients with worse (better) prognosis.
- **Risk assessment**, leading to preventive interventions for those at sufficient risk.
- **Prediction** of safety or efficacy of a specific therapy to aid benefit/risk assessment in individual patients (e.g., predict response, predict SAE, monitor response to adjust schedule or dose or to discontinue).

From AACR-FDA-NCI Cancer Biomarkers Collaborative, Biomarker Assay Validation Subcommittee

Examples of Diagnostic Tests

- Hematology analyzers
- Blood pressure tests
- Cancer screening tests

Validation Studies

- The validation studies for diagnostic devices include analytical validation studies and clinical validation studies.
- Analytical validation establishes performance characteristics including
  - Detection limits (Limit of Blank, Limit of detection, Limit of Quantitation)
  - Precision
  - Accuracy
  - Linear range
  - Reference range
  - Stability
  - Matrix effect
  - …
- Clinical validation evaluates a device's ability to demonstrate clinically meaningful results such as diagnosing and monitoring the target clinical condition, or predicting the onset of a future condition or a treatment response.

Study Design Considerations

- The study design and statistical analysis should support the device intended use.
- Appropriate inclusion/exclusion criteria to ensure that the study population can represent the intended use population.
- Pre-specified statistical analysis plan and study hypothesis/performance goal/acceptance criteria.
- The performance goal and study acceptance criteria should be clinically justified.

Independent Validation

- Validation dataset should be completely independent of derivation dataset.
- Randomly splitting a cohort into training and validation sets may not be adequate:
  - Difficult to document that data leakage did not occur
  - Test may be trained to measurement errors peculiar to that particular dataset, leading to unreproducible performance.

Sources of Bias

- Selection bias (e.g. Study population does not represent intended use population)
- Verification bias (e.g. test positive samples may not take additional confirmative testing)
- Imperfect Reference Standard Bias
- Lead-time bias. (Early Dx doesn't necessarily mean longer life.)
- Length-time bias. (In cross-sectional studies of prevalent cases, enrolled subjects tend to have slower growing disease.)
- Hawthorne bias. (In open label trials, awareness of interventional arm can change behavior of physician, patient.)
- Missing data
- ……

Commonly Used Statistical Analysis Methods

Analytical Performance

- Regression analysis (Linear regression/Deming regression/Passing-Bablok regression/Rank based regression)
- Sensitivity/Specificity
- Positive and negative agreement (PPA and NPA)
- Positive and Negative Predictive Value (PPV and NPV)
- Positive and Negative Likelihood ratio (PLR and NLR)
- Receiver operating characteristic (ROC) and area under the curve (AUC)
- Variance component analysis (Precision)
- Average Positive Agreement and Average Negative Agreement (APA and ANA)
- Agreement Rate with Expected Outcome/Majority Call Outcome (Qualitative/semi-quantitative assay)
- …

Clinical Performance

- Survival analysis
- Categorical data analysis
- Longitudinal data analysis
- …

Variance Component Analysis

- Variance component analysis is commonly used to evaluate the precision (repeatability and reproducibility) of diagnostic devices.
- The study design will evaluate different variation factors (e.g. site, lot, instrument, reader, etc.) that can affect the device outcome. The variation factors may be either crossed or nested within other factors.
- Statistical analysis such as mixed model ANOVA are commonly applied to estimate the variance components.
- Results including mean, standard deviation (SD) and percent coefficient of variation (%CV) are commonly reported.
- Example output:

Sample	N	Mean	Within-Day (SD, %CV)	Between-Day (SD, %CV)	Between-Operator (SD, %CV)	Between-Lot (SD, %CV)	Between-Site (SD, %CV)	Total (SD, %CV)
1	100	50	(0.5, 1%)	(0.5, 1%)	(0.25, 0.5%)	(0.75, 1.5%)	(1.0, 2%)	(1.75, 3.5%)

Regression Analysis

- Linear regression analysis (Simple linear regression, Deming regression, Passing-Bablok regression) is commonly performed to evaluate agreement between two methods for measuring the same measurand.

$$Y_i = \beta_0 + \beta_1 X_i,$$

- Where Y are the test device results, X are the comparator device results.
- Under good agreement, the regression intercept should be close to 0. The regression slope should be close to 1. The predicted bias at medical decision points should be close to 0.
- The study hypothesis is usually

$$H0: |\beta_0| > \delta; H1: |\beta_0| \le \delta$$

$$H0: |\beta_1 - 1| > \eta; H1: |\beta_1 - 1| \le \eta$$

$$H0: |Bias| > \lambda; H1: |Bias| \le \lambda,$$

- Where δ, η, and λ are clinically acceptable equivalence margin;

$$Bias = \beta_0 + (\beta_1 - 1)X$$

- Because there may measurement errors for both Y and X, the regression analysis may need to account for measurement errors for both measurements.

Sensitivity/Specificity, Positive and Negative Percent Agreement (1)

	Diseased	No-Diseased
Test Positive	a	b
Test Negative	c	d

- Sensitivity=a/(a+c): the proportion of subjects that are test positive among those with the disease.
- Specificity=d/(b+d): the proportion of subjects that are test negative among those without the disease.
- Sensitivity and specificity evaluates the diagnostic accuracy of a test. We should always look at sensitivity and specificity together.
- If the true disease status is unknown and we compare the device under evaluation to a comparator method, positive percent agreement (PPA) and negative percent agreement (NPA) will be reported. PPA and NPA evaluate agreement instead of accuracy.

Sensitivity/Specificity, Positive and Negative Percent Agreement (2)

Common statistical issues that may arise include

- Paired data: if a study involves testing all sample/subjects using 2 or more devices and a reference standard (truth). The comparison of the accuracy needs to consider paired study design. (ref. NEWCOMBE; 1998; STATISTICS IN MEDICINE)
- Correlated/clustered data: Statistical analysis need to account for possible correlations. (e.g. random effects model, bootstrap method, etc.)
- Verification bias adjustment: If only a subset of negative or positive samples took the confirmative reference testing, verification adjusted estimates should be used so that the performance estimates are not biased.
- Discrepant resolution: Either avoided or appropriate sampling method should be used for resolution.

Bridging Analysis (1)

- A companion diagnostic device (CDx) is “one that provides information that is essential for the safe and effective use of a corresponding therapeutic product”.
- When we enroll subjects to evaluate the therapeutic product, a clinical trial assay (CTA) instead of CDx may be used for patient enrollment in the clinical trial. As a result, the efficacy of the therapeutic product may only be available for the CTA positive population.
- In order to bridge the drug efficacy from CTA positive subject to CDx intended use population, a concordance study between CTA and CDx may be performed.

	CTA=+	CTA=-	Total
CDx=+	a	b	a+b
CDx=-	c	d	c+d
Total	a+c	b+d	a+b+c+d

- PPA=a/(a+c); NPA=d/(b+d);
- PPV=a/(a+b); NPV=d/(c+d);
- If the concordance is enriched to get more positive samples, the calculation PPV and NPV needs to adjust for the true prevalence of condition in the intended us population.

Bridging Analysis (2)

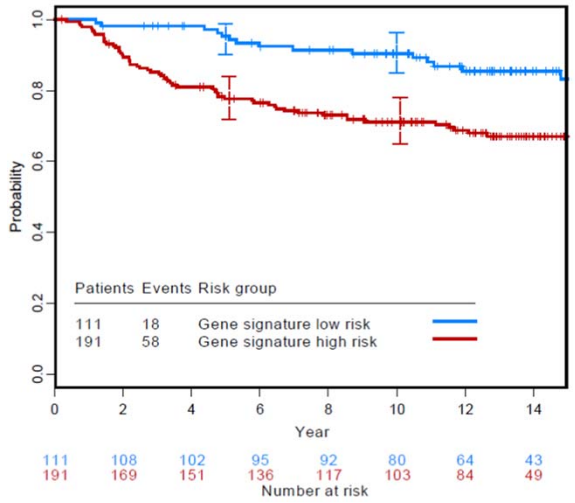
- Suppose the efficacy of a treatment is response rate (R), the efficacy for the CDx positive population can calculated as

$$P(R=1|CDx=+)=P(R=1|CDx=+,CTA=+)*PPV+P(R=1|CDx=+,CTA=-)*(1-PPV)$$

- To support effectiveness of a CDx, generally the bridging study and analysis should demonstrate that drug efficacy is maintained for the CDx positive subjects compared to CTA positive subjects.
- More recommendation about study designs and statistical analysis are available at Meijuan Li (2015) Statistical Consideration and Challenges in Bridging Study of Personalized Medicine, Journal of Biopharmaceutical Statistics, 25:3, 397-407

Clinical Outcome

- The clinical validity of a diagnostic test can also be evaluated based on clinical outcome.
- For example, MammaPrint (K062694; a prognostic assay) was evaluated based on probability of distant metastasis at different follow up times



5-year:
Low risk group: 0.95 (0.91-0.99)
High risk group: 0.78 (0.72-0.84)

10-year:
Low risk group: 0.90 (0.85-0.96)
High risk group:0.71 (0.65-0.78)

CLSI Guidances for Evaluating Diagnostic Devices

- EP05-A3 Precision
- EP06-A Linearity
- EP07-A2 Interference Testing
- EP09-A2 Systematic difference (bias)
- EP12-A2 Qualitative Test Performance
- EP14-A3 Commutability
- EP17-A2 LoB, LoD and LoQ
- EP21-A Total error (accuracy)
- EP25-A Stability of reagents
- EP28-A3c Reference intervals
- MM17-A Multiplex tests
- EP32-R Traceability

FDA Guidances

- Design Considerations for Pivotal Clinical Investigations of Medical Devices *(2013)*
- Statistical Guidance for Reporting Results from Studies Evaluating Diagnostic Tests *(2007)*
- Adaptive Designs for Medical Device Clinical Studies (2016)
- FDA Guidance for Use of Bayesian Statistics (2010)
- Assay Migration Studies for In Vitro Diagnostic Devices (2013)
- ……

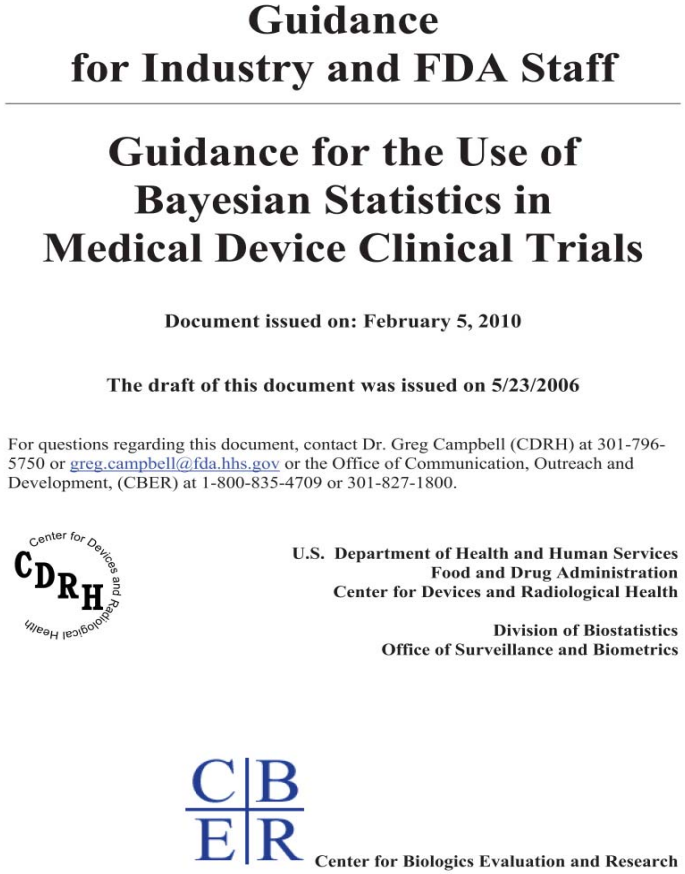
Applications of Bayesian Methods in Medical Device Clinical Studies

- Bayesian study designs/methodologies have been used to support various medical device clinical studies.
  - Incorporation of prior information in the treatment or in the control (or both)
  - Bayesian adaptive designs
  - Facilitate stopping decisions by company and DMCs (interim analysis)
  - Confirm surrogate endpoint (prediction)
  - Evaluate changes in a trial or device during the conduct of the trial
  - assessing equivalence (non-inferiority)
  - …

Examples of Bayesian Device Studies

- BRYAN Cervical Disc by Medtronic Sofamor-Danek(approved; P060023 SSED available on Web)
  - Bayesian statistical methods were used to determine whether the investigational device is non-inferior to the control with respect to the overall success rate.
- AngelMed Guardian System (P150009 Advisory Committees Meeting Materials; available on Web)
  - Bayesian adaptive design was selected so that sample size could be dynamically determined during the course of the trial. Posterior probability was used to assess the level of evidence in support of a hypothesis.

FDA Guidance for Use of Bayesian Statistics



Guidance
for Industry and FDA Staff

Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials

Document issued on: February 5, 2010

The draft of this document was issued on 5/23/2006

For questions regarding this document, contact Dr. Greg Campbell (CDRH) at 301-796-5750 or greg.campbell@fda.hhs.gov or the Office of Communication, Outreach and Development, (CBER) at 1-800-835-4709 or 301-827-1800.

Center for Devices and Radiological Health
CDRH

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Devices and Radiological Health

Division of Biostatistics
Office of Surveillance and Biometrics

CBER
Center for Biologics Evaluation and Research

Acknowledgements

- Gene Pennello, Ph.D.
- Division of Biostatistics