

Data Analysis: Software and Statistics

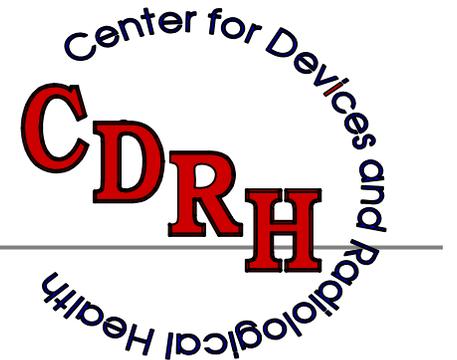
Estelle Russek-Cohen

Office of Surveillance and Biometrics

Division of Biostatistics

Center for Devices and Radiological Health

U.S. Food and Drug Administration



Caveat

- Views expressed are those of the author.

Outline

- Software

 - Why is Software different?

 - Documentation

 - Level of Concern

 - Validation and Verification

- Valid Scientific Evidence

- Statistics

 - Study designs: Non-clinical/Clinical

 - Confirmatory analyses

Software is different

- Software quality: Primarily a design issue
- Complexity
- Dormant latent defects
- Software is easy to change
- Difficult to control changes
- Significance of changes
- Structured development process plus testing

Software-Related Premarket Guidance Documents

- **Guidance for the Content of Premarket Submissions for Software Contained in Medical Devices (May 2005)**
- **Guidance for Off-the-Shelf Software Use in Medical Devices**
- **The New 510(k) Paradigm: Alternate Approaches to Demonstrating Substantial Equivalence in Premarket Notifications**
- **General Principles of Software Validation**
- **Guidance for Industry: Cybersecurity for Networked Medical Devices Containing Off-the-Shelf (OTS) Software**

Other available resources

- Standards and Guidelines

 - ISO/IEC 12207

 - IEEE/EIA 12207

 - ISO 9001 and 9000-3

 - IEEE Software Standards Compendium

 - ANSI/AAMI SW68:2001

See last slide....

Level of Concern

- Separate from Device Classification
- Concern defined by the sponsor
- Level of concern determines recommendations for documentation

Levels of Concern (cont'd):

- **Major**: malfunction may cause Death or Serious Injury
- **Moderate**: malfunction may cause injury other than serious
- **Minor**: malfunction unlikely to cause injury
- Injuries may be to patients and / or operators

Principles

- Defect Prevention - **better than finding & fixing defects**
- Software Life Cycle - **defined activities, tasks, and documentation**
- Time and Effort - **validation throughout life cycle**
- Plans - **“what”**
- Procedures - **“how”**

Principles (Cont.)

- Requirements - **documented baseline for V & V**
- Testing - **mix of static analyses and dynamic testing**
- Test Coverage - **commensurate with risk and complexity**
- After Changes - **impact analysis & regression testing**
- Independence of review - **preferred where possible**
- Real World - **flexibility, but also responsibility**

Validation, Verification and Testing

- Test with users involving realistic use, objective performance information and subjective evaluation by users
- Are alarms, messages, etc sufficient for user to use device correctly?
- Describe V, V & T activities at unit level, integration and system level.
- Testing results including Hazard analysis and software functional requirements.

Off The Shelf Software

- Device manufacturer still bears responsibility for safe and effective performance of device
- Provide basic documentation of software
- Perform device and OTS software hazard analysis
- If needed, manufacturer needs to identify how it deals with hazard mitigation
- See guidance or contact FDA

Software Validation Required

- For device software

 - § 820.30(g) “... design validation shall include software validation and risk analysis where appropriate...”

- For automation of production processes or automation of the quality system

 - § 820.70(i) “... the [device] manufacturer shall validate computer software for its intended use according to an established protocol. All software changes shall be validated before approval and issuance. These validation activities and results shall be documented.”

Software updates

- If software changes alter the safety and effectiveness of the device
 - Expect to contact FDA
- Formal mechanisms:
 - New 510(k)
 - PMA Supplement
- May need more data

Useful contacts on software:

- Joseph Jorgens
Team Leader for Software Evaluation
Division of Electrical and Software
Evaluation
Joseph.Jorgens@fda.hhs.gov
- James Callaghan
Office of In Vitro Diagnostics
James.Callaghan@fda.hhs.gov

Valid Scientific Evidence

- Studies presented by the sponsor should support
 - intended use of the device
 - intended use population
 - intended use setting
- Multiple intended uses may require multiple studies.

Academic vs FDA research

- Academia:

 - Tight lab control of extraneous sources of variability...isolate effects of interest

- FDA

 - Looks for a device that can operate in many situations in spite of extraneous sources of variability

Academia vs FDA Research

- Exploratory methods are useful in academia and perhaps in product development
- FDA wants to see hypothesis driven confirmatory studies for approval of devices
- One or more measures of performance must be provided.

21CFR809.10 Device Labeling

- “Explanation of the procedure for calculating the unknown”
- “Specific performance characteristics including accuracy, specificity, precision and sensitivity”
- Sensitivity= $P(\text{test}=+|\text{patient}=+)$
- Specificity= $P(\text{test}=-|\text{patient}=-)$

Microarrays at FDA

- Roche Amplichip Genotyping Array
Cleared December 2004
- SNPs of two Cytochrome P450 genes
- Affymetrix platform
- Pair of deNovo 510(k)s
- Special controls guidances for each on-line

Microarrays

- Lots of probes
- Complex normalization schemes
- Function to combine many measurements to get to a classification or diagnosis

Conventional stat algorithms

Machine learning algorithms

- How to evaluate performance?

Some Cautions: Academia vs FDA

- Use normalizations within an **experiment**
vs
- Normalization within an **array**
- Two color arrays:
Need reference material that stays stable over:
experiment vs multiple labs over time.
- May use **one** lab/site in academia
- We often ask for **three** labs and/or clinical sites

Training vs Validation

- Training often involves looking at many traits to get down to a few
- Training often involves looking at different normalizations and classifiers to optimize performance
- Validation on the training set leads to overly optimistic view of performance
- Cross-validation methods are not sufficient

Journal article

- FDA Perspectives on potential microarray-based clinical diagnostics

Tezek et al (January 2006)

Human Genomics 2(4):236-243

Performance Validation

- Sample must be reflective of intended use
- Classifier is fixed as it will go to market
- Use validation set that is independent of training set –ideally new study sites
- Evaluate sensitivity and specificity if a **true diagnosis** is known
- Can evaluate agreement with an already marketed device for same use.

“Truth” or True diagnosis

Some examples:

- Genotyping: bidirectional sequencing
- Cancer: biopsy and pathology report
- Could be the result of a panel of experts all having access to the complete case records

- Check with FDA before study

Confirmatory analysis

- Sensitivity exceeds a threshold
Lower 95% confidence bound exceeds
e.g. **90%**
- Specificity exceeds a threshold
Lower 95% confidence bound exceeds
e.g. **90%**
- Use exact confidence intervals for binomial proportions
- **Check with FDA concerning performance threshold**

Quantitative assays: single measure

- Measuring whole body radiation exposure:
Demonstrate comparability to a reference standard over measurement range
- Common approaches: a regression method with reference= X , your device= Y .
- Report confidence intervals; systematic bias
- Check with FDA on choice of reference standard

Accounting for the data

- Clearly defined protocol
- Eligibility: Inclusions/Exclusions
- Missing patients and missing specimens
 - Why and how many
- How samples are processed (e.g. FFPE)
- Study limitations

In Vitro Diagnostics

- **Non-clinical studies including:**

- Accuracy and precision

- Over device lots, reagents lots,...

- Reproducibility (3 labs)

- Interference

- Linearity for quantitative assays

- Performance near the cutoff for qualitative assays

-
- Use many of the guidances :

Clinical Laboratory Standards Institute

<http://www.clsi.org>

Evaluation Protocols (e.g. EP5, EP6,)

Molecular Methods

Some **ISO** guidances, some **ASTM** guidances

Banked Tissue/Blood Samples

- Can these get used in a submission?

YES

- Caveats

Samples should be representative of intended use population

Need accounting for missing data, study biases,

Animal Studies

- May be allowed if people studies are unethical
- Need to demonstrate relevance to people
- Study designs should be consistent with quality Veterinary submissions
- Check with FDA first!

Conclusions

- Each device is unique.
- Guidances are helpful.
- Smart sponsors make use of the pre-IDE process prior to doing major studies.
- Documenting the science, the data analysis and the software as part of your submission is important.

Finding out more on the web

- <http://www.fda.gov>
- Click on Medical devices
- Click on Device advice
 -search for Callaghan (OIVD software)
- Click on guidances
- Click on CDRH databases: list of standards recognized by FDA

Thanks.....

Estelle Russek-Cohen

Estelle.Russek-Cohen@fda.hhs.gov