U.S. FOOD & DRUG ADMINISTRATION

Statistical Computing Challenges at FDA

Paul Schuette, Ph.D.
Scientific Computing Coordinator
Office of Biostatistics
FDA/CDER/OTS

What does the FDA do?

- FDA is responsible for protecting the public health by assuring the safety, efficacy and security of human and veterinary drugs, biological products, medical devices, our nation's food supply, cosmetics, and products that emit radiation.
- FDA regulated products comprise approximately 20%-25% of consumer spending in the United States.

FDA Structure

The US FDA is organized by Centers:

- Center for Biologics Evaluation and Research (CBER)
- Center for Drug Evaluation and Research (CDER)
- Center for Devices and Radiological Health (CDRH)
- Center for Food Safety and Applied Nutrition (CFSAN)
- Center for Tobacco Products (CTP)
- Center for Veterinary Medicine (CVM)
- National Center for Toxicological Research (NCTR)

CDER's Office of Biostatistics

- Over 200 Statisticians

- Vision

The Office of Biostatistics is recognized for excellence in the application and communication of statistical science in drug regulation and development. We play a central role in promoting innovative, science-based, quantitative decision-making throughout the drug development life-cycle.

- Mission

Provide CDER and other internal and external stakeholders with statistical leadership, expertise, and advice to foster the expeditious development of safe and effective drugs and therapeutic biologics for the American people. Protect the public health by applying statistical approaches for monitoring the effectiveness and safety of marketed drugs and therapeutic biologic products.

Hardware

Principal Statistical Computing Environments at FDA:

- Basic Business Laptops: 8-16G RAM, 2-4 cores, 300-500G local storage (hard drives)
- Scientific Workstations: 32G-192G RAM, 2-12 cores, 300G-8T local storage
- FDA High Performance Computing (HPC), Grid Engine architecture, 359 nodes, 3168 cores, large amounts of RAM and storage. Linux OS. Primarily used for research, prototyping and genomic analyses.

Statistical Software Clarifying Statement

"FDA does not require use of any specific software for statistical analyses, and statistical software is not explicitly discussed in Title 21 of the Code of Federal Regulations [e.g. in 21CFR part 11]. However, the software package(s) used for statistical analyses should be fully documented in the submission, including version and build identification."

Statistical Software Clarifying Statement, continued

"As noted in the FDA guidance*, E9 Statistical Principles for Clinical Trials*, `The computer software used for data management and statistical analysis should be reliable, and documentation of appropriate software testing procedures should be available.' Sponsors are encouraged to consult with FDA review teams and especially with FDA statisticians regarding the choice and suitability of statistical software packages at an early stage in the product development process. "

http://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/ucm445917.htm

Statistical Software used at FDA

- SAS
- R, RStudio
- Stata
- Minitab
- Splus

- MATLAB
- Mathematica

- StatXact, LogXact
- EAST
- nQuery+nTerim
- ADDPLAN, FACTS
- PASS
- Comprehensive Meta Analysis
- OpenBUGS, JAGS, Stan
- Stat/Transfer

Statistical Software Challenges

Client-Server Software can be problematic:

- Requires approval by the Office of Information Management and Technology (OIMT) at FDA.
- Time consuming, nontrivial approval process
- Expensive, requires ~60K to pay for security review by OIMT contractors, etc.
- Annual costs still substantial.
- OIMT requires Development, Test, Preproduction and Production Environments for most applications.
- SAS Server due to launch in 2017, requested in 2012.
- R Shiny Server pilot project requested for 2017.

Some Solutions

- Co-operative statistical software purchasing among major centers (CBER, CDER, CDRH).
- Blanket Purchasing Agreement for Scientific Workstations
- Resource sharing, such as the FDA HPC, but shared governance issues exist
- Open Source Software, particularly R, facilitates work with external collaborators:
  - on Broad Agency Agreements, BAA.
  - Cooperative Research and Development Agreements (CRADA).
  - openFDA initiative, externally hosted

Data Challenges

Data Integrity and Quality can be a challenge across centers.

- Sponsors submit data, programs, and analyses to FDA.
- FDA is a member of the **C**linical **D**ata **I**nterchange **S**tandards **C**onsortium (CDISC).
- CDISC standards help but ….
  - Multiple interpretations of CDISC standards
  - Compliance with standards, compliance checks can be an issue
  - ADaM standards are still undergoing development
- CluePoints CRADA for detecting and identifying data anomalies
- Future challenges: genomic data and electronic health records, Cancer Moonshot, Precision Medical Initiative

Providing Regulatory Submissions In Electronic Format — Standardized Study Data

Guidance for Industry

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

December 2014
Electronic Submissions

STUDY DATA TECHNICAL CONFORMANCE GUIDE

Technical Specifications Document

This Document is incorporated by reference into the following Guidance Document(s):

Guidance for Industry *Providing Regulatory Submissions in Electronic Format – Standardized Study Data*

For questions regarding this technical specifications document, contact CDER at cder-edata@fda.hhs.gov or CBER at cber.cdisc@fda.hhs.gov

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

July 2016

Transparency, Repeatability and Reproducibility

- FDA reviewers must evaluate sponsors' trial designs, analyses, and claims.
- Sponsors may use multiple internal macros to prepare and analyze data.
- Sponsors may outsource trial design, data collection, data management, and analyses to **C**ontract **R**esearch **O**rganizations (CROs)
- FDA reviewers may not always agree with sponsor's results

Some Recommendations

- Specify version and build of statistical software programs.
- Specify all packages and components used, including version and build.
- Specify operating system and hardware.
- Specify and document any changes or tweaks to configuration files.
- If macros are used, specify dependencies and necessary file structures.
- Flowchart workflow necessary to produce analyses and analysis data sets.

Additional Recommendations

- Follow Good Programing Practices (GPP)
  - Thoroughly document programs.
  - See PhUse wiki on GPP

    http://www.phusewiki.org/wiki/index.php?title=Good_Programming_Practice#What_is_GPP
  - SAS® Programming Guidelines, http://www2.sas.com/proceedings/sugi31/123-31.pdf

Personnel

- Statistical Reviews are conducted by FDA statisticians, who typically have a background in statistical methodology and analysis, rather than database management and programming.
- Most reviewers have responsibility for multiple reviews (INDs, NDAs, BLAs) at any given time.
- CDER is adding Statistical Analysts (1530 series) to help support statistical computing needs of reviewers in the Office of Biostatistics.

Historical Note

1971 President's Commission Report by Daniel, Tufte and Kadane:

"The lack of programming support has also led to some inefficient efforts at statistical programming by Ph.D. level staff."

"The Bureau of Drugs needs much greater support with respect to scientific programming: data processing facilities of an order of magnitude greater and better than those now existing are predictably needed by the Bureau in the long run."

Remarks and Conclusions

1. Statistical Computing at FDA is improving, but major challenges remain.
2. With more complex designs (Bayesian and adaptive) and analyses (tipping point, Bayesian analyses) there is more demand for statistical computing.
3. "In a world when anything is possible because you have so much data, how do you figure out who has done the math right?" FDA Commissioner Robert Califf
4. Transparency, repeatability and reproducibility in statistical computing are critical.
5. Both sponsors and regulators would benefit by adopting Good Programming Practices.