Statistical Computing Environments in CDER

Paul Schuette email:Paul.Schuette@fda.hhs.gov

Office of Biostatistics, Immediate Office, FDA/CDER/OTS

Disclaimer

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

Abstract

Statistical reviewers in CDER's Office of Biostatistics (OB) may use several possible statistical computing environments (SCEs), depending on their needs. Hardware available to reviewers includes standard business laptops, scientific workstations and FDA's high performance computing (HPC) environment. Reviewers may use a variety of statistical software packages such as SAS, R, and Stata as well as more specialized software packages such as EAST, StatXact, nQuery, PASS, MATLAB and Mathematica. We describe a pilot program to stand up an enterprise statistical computing environment on a server. Finally, we describe an initiative to add dedicated statistical analysts to assist reviewers with statistical computing and programming needs. Understanding the performance capabilities of CDER's statistical computing environments may help sponsors prepare and plan for submissions that can be readily reviewed by OB statisticians.

Data

How do reviewers get data from sponsors?

1. Sponsors submit data to the Electronic Submissions Gateway (ESG)
2. Request number, Conformance checks
3. Data loaded into the Electronic Document Room (EDR)

Reviewers can then pull data from the EDR for analysis, as well as requesting additional checks against CDISC standards (Jumpstart). Reviewers cannot change the data in the EDR.

SCE Hardware

Statistical Reviewers in CDER's Office of Biostatistics (OB) may use one of three statistical computing environments for their analyses:

1. Basic Laptops. 8-16G RAM, 300G hard drive, 8 cores. Windows 7 OS, Computationally limited.
2. Scientific Workstations and Scientific Laptops. 25 shared workstations, 53 Scientific Laptops, 24-192G RAM, 240G SSD to 2T hard drives, 8-24 cores, Windows 7 OS.
3. FDA High Performance Computing (HPC), Grid Engine architecture, 359 nodes, 3168 cores, large amounts of RAM and storage. Linux OS. Helpful to know standard Unix tools such as bash and vi.

SCE Software

Statistical Reviewers may use a variety for statistical software packages for their analyses. General statistical packages for Windows:

SAS, R, Splus, Stata, Minitab, Statistica.

For the HPC, SAS and R are commonly used.

SCE Software Continued

Specialized statistical packages include:

StatXact, LogXact, EAST, nQuery+nTerim, ADDPLAN, PASS, StatTransfer, RStudio, Comprehensive Meta Analysis, OpenMetaAnalyst OpenBUGS, JAGS, Stan

General scientific computing package include:

MATLAB, Mathematica, Python, Fortran, C++

FDA reviewers are able to use a wide range of statistical and scientific computing software packages.

Statistical Software Clarifying Statement

FDA is software agnostic:

"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. ...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." See http://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/ucm445917.htm for details.

FDA Initiatives

- **Server based SAS.** Initiated with the FDA Scientific Computing Board and the Office of Information Management and Technology (OIMT). Initially intended to address needs of occasional users, and to pilot using analytic products such as Enterprise Miner and Visual Analytics in a secure server environment.
- **Statistical Analysts Positions.** The Office of Biostatistics has begun recruiting for Statistical Analysts positions (1530 series). Statistical analysts will assist with programming and support needs of OB. Specifically, "Develops software using the appropriate statistical programming packages (e.g., R, JAGS, SAS) for OB statistical reviewers to support programming intensive review-related activities, such as, sensitivity analysis, Bayesian approaches, clinical trials modeling, genomic studies, psychometric Clinical Outcome Assessment (COA) validation and simulation."

Additionally, "Designs and programs data checks, as needed, to facilitate data quality and integrity. Inspects data received from sponsors and other sources for problematic entries and missing values. Evaluates and reports quality of submitted clinical trials data and the data from large safety databases. Monitors the quality of the implementation of data standards used in applicant submissions."

Candidates at the masters level or above are sought, and position descriptions have been posted to USAJobs.