R for Clinical Trial Reporting: Reproducible Research, Quality and Validation

Frank E Harrell Jr

Department of Biostatistics, Vanderbilt University School of Medicine

useR! 2007 Conference 10 Aug 2007

Slides and Code at `http://biostat.mc.vanderbilt.edu/Rreport`

Outline

1. Software Quality and Validation
   - Quality and Error Sources
   - What is Called "Validation" and What Should it Be?
   - Example of a Comprehensive Analysis Validation
2. High-Level Tools for Reproducible Analysis and Reporting
   - Background
   - Tools
   - Statistical Methods
   - Example

Quality and Error Sources: Example

- Since around 1967 SAS treats `NA` as $-\infty$
- Key analysis from Duke U. published in NEJM used

```
IF stroke_time < follow_up_time THEN stroke=1;
```

- Patients having missing `stroke_time` categorized as having stroke (2° endpoint, primary was death)
- Never corrected

Some Sources of Errors

- Original information source
- Data entry and OCR
- Derived variables
- Data management and storage
- Data import/conversion package
- Data manipulation and analysis file creation
- Statistical package/system
- User analysis code
- Transcription of results into report
- Error in insertion or typesetting results
- Interpretation of results

Most Common Errors Involving Analysts

- Derived variables
- Data manipulation and analysis file creation
- Errors in user analysis code

What is "Validation"?

- Traditionally it involves validating general statistical packages through
  - code inspection
  - test cases
  - simulation
- Such validation cannot envision all possible combinations of options / analyses, or all possible data configurations

What Should Validation Really Emphasize?

Validation of **analyses**

- Entire process of analysis file creation, analysis, graphics
- Resources seldom available for first part
  - Analysis file creation tested interactively, merge datasets and derive variables two ways, etc.
- Validation is not static is per-analysis
- For pivotal analyses, compare results (point estimates, confidence intervals, *P*-values) with those from another package
- For checking R calculations, ideal independent and highly programmable package is probably Stata

A Comprehensive Validation

- Statistical Center (SC) at Vanderbilt does not use SAS for any aspect of data processing or analysis except sometimes to export data from SAS
- Sponsor uses SAS for all data manipulation, derived variables, analysis
- SC created dummy randomization and created an unblinded study report; sent to sponsor
- Sponsor recreated all pivotal calculations
- Worked to obtain exact agreement
- Biggest challenge: getting exactly same study samples (e.g., "efficacy population")

High Level Tools: Purpose

- Data Monitoring Committees
- Enhance safety and risk/benefit review by DMC
- Methods useful for general RCT reports
- Provide efficient and state-of-the-art statistical reporting
- High-quality graphics (a la Bill Cleveland) and tables
- Hard copy and on-screen review

Problems to Solve

- Reproducible research: no transcription of results
- Repeated reports, main changes are updates to data
- Many response variables and repeated measurements
- Non-normality of data (especially clinical chemistry)
- Dropouts and missing data
- Graphical methods for judging differences in point estimates

Tools Needed

- Batch mode capability (scripting)
- Fine control (graphics, tables, text)
- High-level, flexible statistical language
  - graphics
  - statistical analysis
  - easy to implement new functions
  - functions are data-sensitive (unlike macros)
  - advanced tables

Tools Needed, *cont.*

- Document processing (typesetting)
  - easy handling of Greek letters, subscripts, superscripts, font changes
  - no cut and paste
  - easy inclusion of chunks of text, tables, graphics
  - automatic cross-referencing and hyperlinking
  - let software worry about formatting details

Tools Selected and Developed

- R: open source statistical language
- LaTeX
- Hmisc package
  - advanced table making
  - `latex` functions to convert S objects to LaTeX code
  - graphics
  - Lan-DeMets sequential monitoring stopping bands
- Design package for survival curve plotting

Tools, *cont.*

- New series (`rreport` package) of higher-level report generation functions
  - `completenessReport`, `accrualReport`, `baselineReport`, `mixedvarReport`, `repVarclus`, `complianceReport`, `dropoutReport`, `aeReport`, `labReport`, `publishPdf`, `mockTable` functions
  - uses data attributes (value levels, variable labels, units)
  - generates all tables, graphs, figure and table captions
  - unified mapping of treatments to line types, with graphical legends in text captions
  - generates some sentences
  - conditional inclusion of certain graphics and sentences

Tools, *cont.*

- All non-graphical output files are LaTeX
- Generates all LaTeX `\includegraphics` calls
- Simultaneously generates open and closed meeting components
- User writes calls to modular functions, study-specific text
- `pdf` file created directly by `pdflatex`
- `hyperref` style used for automatic hyperlinking
- `publishPdf` function copies reports to secure web server, creates `html index` file for them, and e-mails committee members and assistants URLs, access IDs, and (sep. e-mail) passwords

Graphical Method for Interpreting Differences

- Confidence limits have more information that P-values
- Graphs showing CLs for multiple treatment groups are busy
- Confidence interval for difference in two parameters not directly obtainable from individual confidence intervals
- Best to show individual estimates and include a separate panel to show difference and its CLs
- Compromise: draw half-width of CL centered at midpoint of two estimates

$$\begin{aligned} \frac{\bar{Y}_1 - \bar{Y}_2}{se} &> z \\ \bar{Y}_1 - \bar{Y}_2 &> z \times se \\ \text{Width of CL} &= 2 \times z \times se \end{aligned}$$

Example

- Data from an actual clinical trial, contributed from a pharmaceutical company
- Not included in example report:
  - efficacy analysis
  - study design
  - data monitoring plan
  - summary of previous closed reports
  - interpretation
  - protocol changes
  - screening
  - eligibility
  - waiting time until treatment commencement
- See Ellenberg, Fleming, DeMets: *Data Monitoring Committees in Clinical Trials*, 2002

R for Clinical Trial Reporting: Reproducible Research, Quality and Validation

Frank E Harrell Jr
Vanderbilt University

Reliability of analysis software is of paramount importance in clinical and pharmaceutical research. Classical software "validation" has little to do with quality, as most errors are committed when deriving variables, manipulating and analyzing data. Validation should be directed towards checking the analysis at hand.

The methods often used for generating statistical reports for clinical trials have a number of drawbacks. The most commonly used statistical software packages require users to specify somewhat tedious low-level commands, and the resulting tabular and graphical output are not optimal. Too often, statisticians still overuse tabular reports even though most consumers of the reports would rather review graphics. And in an era in which reproducible research is starting to become popular, most statisticians still engage in some level of manual intervention, such as insertion of calculated values in sentences. These issues are particularly important in reporting for data monitoring committees.

This talk will describe an approach that uses free open-source software (R and LaTeX) to produce advanced tables and graphics using a very high-level language. The component tables and graphics are automatically assembled and indexed by LaTeX, resulting in an Adobe Acrobat PDF file with hyperlinks for easy navigation. Example open- and closed-session DMC reports will be shown, which includes tables and graphics describing data completeness, subject accrual, baseline variables, compliance, dropouts, adverse events, and lab data. Some issues in statistical graphics will be discussed, such as a way to depict confidence limits for differences between treatments in graphs that show individual treatment responses.