

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

R for Clinical
Trial
Reporting

Vanderbilt
Biostatistics

Software
Quality and
Validation

Quality and
Error Sources

What is Called
“Validation” and
What Should it
Be?

Example of a
Comprehensive
Analysis
Validation

High-Level
Tools for
Reproducible
Analysis and
Reporting

Background

Tools

Mock Tables

Statistical
Methods

Example

- 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
 - Mock Tables
 - Statistical Methods
 - Example

Quality and Error Sources: Example

R for Clinical
Trial
Reporting

Vanderbilt
Biostatistics

Software
Quality and
Validation

Quality and
Error Sources

What is Called
"Validation" and
What Should it
Be?

Example of a
Comprehensive
Analysis
Validation

High-Level
Tools for
Reproducible
Analysis and
Reporting

Background
Tools

Mock Tables
Statistical
Methods
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^o endpoint, primary was death)
- Never corrected

Some Sources of Errors

R for Clinical
Trial
Reporting

Vanderbilt
Biostatistics

Software
Quality and
Validation

Quality and
Error Sources

What is Called
"Validation" and
What Should it
Be?

Example of a
Comprehensive
Analysis
Validation

High-Level
Tools for
Reproducible
Analysis and
Reporting

Background
Tools
Mock Tables
Statistical
Methods
Example

- 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

R for Clinical
Trial
Reporting

Vanderbilt
Biostatistics

Software
Quality and
Validation

Quality and
Error Sources

What is Called
"Validation" and
What Should it
Be?

Example of a
Comprehensive
Analysis
Validation

High-Level
Tools for
Reproducible
Analysis and
Reporting

Background
Tools

Mock Tables
Statistical
Methods
Example

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

What is “Validation”?

R for Clinical
Trial
Reporting

Vanderbilt
Biostatistics

Software
Quality and
Validation

Quality and
Error Sources

What is Called
“Validation” and
What Should it
Be?

Example of a
Comprehensive
Analysis
Validation

High-Level
Tools for
Reproducible
Analysis and
Reporting

Background
Tools

Mock Tables
Statistical
Methods
Example

- 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?

R for Clinical
Trial
Reporting

Vanderbilt
Biostatistics

Software
Quality and
Validation

Quality and
Error Sources

What is Called
"Validation" and
What Should it
Be?

Example of a
Comprehensive
Analysis
Validation

High-Level
Tools for
Reproducible
Analysis and
Reporting

Background
Tools

Mock Tables
Statistical
Methods
Example

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

R for Clinical
Trial
Reporting

Vanderbilt
Biostatistics

Software
Quality and
Validation

Quality and
Error Sources

What is Called
"Validation" and
What Should it
Be?

Example of a
Comprehensive
Analysis
Validation

High-Level
Tools for
Reproducible
Analysis and
Reporting

Background
Tools

Mock Tables
Statistical
Methods
Example

- 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

R for Clinical
Trial
Reporting

Vanderbilt
Biostatistics

Software
Quality and
Validation

Quality and
Error Sources

What is Called
"Validation" and
What Should it
Be?

Example of a
Comprehensive
Analysis
Validation

High-Level
Tools for
Reproducible
Analysis and
Reporting

Background
Tools

Mock Tables
Statistical
Methods
Example

- 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

R for Clinical
Trial
Reporting

Vanderbilt
Biostatistics

Software
Quality and
Validation

Quality and
Error Sources
What is Called
"Validation" and
What Should it
Be?

Example of a
Comprehensive
Analysis
Validation

High-Level
Tools for
Reproducible
Analysis and
Reporting

Background
Tools
Mock Tables
Statistical
Methods
Example

- 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

R for Clinical
Trial
Reporting

Vanderbilt
Biostatistics

Software
Quality and
Validation
Quality and
Error Sources
What is Called
"Validation" and
What Should it
Be?
Example of a
Comprehensive
Analysis
Validation

High-Level
Tools for
Reproducible
Analysis and
Reporting
Background
Tools
Mock Tables
Statistical
Methods
Example

- 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.*

R for Clinical
Trial
Reporting

Vanderbilt
Biostatistics

Software
Quality and
Validation

Quality and
Error Sources

What is Called
"Validation" and
What Should it
Be?

Example of a
Comprehensive
Analysis
Validation

High-Level
Tools for
Reproducible
Analysis and
Reporting

Background

Tools

Mock Tables

Statistical
Methods

Example

- 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 for Clinical
Trial
Reporting

Vanderbilt
Biostatistics

Software
Quality and
Validation

Quality and
Error Sources

What is Called
"Validation" and
What Should it
Be?

Example of a
Comprehensive
Analysis
Validation

High-Level
Tools for
Reproducible
Analysis and
Reporting

Background

Tools

Mock Tables

Statistical
Methods

Example

- 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

- 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.*

R for Clinical
Trial
Reporting

Vanderbilt
Biostatistics

Software
Quality and
Validation

Quality and
Error Sources

What is Called
"Validation" and
What Should it
Be?

Example of a
Comprehensive
Analysis
Validation

High-Level
Tools for
Reproducible
Analysis and
Reporting

Background

Tools

Mock Tables

Statistical
Methods

Example

- 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

Concise Language for Mock Tables for SAPs

R for Clinical
Trial
Reporting

Vanderbilt
Biostatistics

Software
Quality and
Validation

Quality and
Error Sources
What is Called
"Validation" and
What Should it
Be?

Example of a
Comprehensive
Analysis
Validation

High-Level
Tools for
Reproducible
Analysis and
Reporting

Background
Tools
Mock Tables
Statistical
Methods
Example

S commands:

```
mockTable(list(  
  treatments = c('Placebo', '0.01 mg/kg', '0.1 mg/kg',  
                '1 mg/kg'),  
  Pvalues    = c('$P$\nANOVA\\tmark', '$P$\nK-W'),  
  CLs       = list('Bootstrap Nonparametric 95\\% Confidence Limits',  
                  'all'),  
  byVar     = c('Age $\leq$ 30d Group',  
                'Age 1-24m Group'),  
  baseListVar = c('Site', 'Subject', 'Age, $d$', 'Dose'),  
  file      = 'mock.tex',  
  keyfile   = 'defs.tex',  
  datafile  = 'mock.dat',  
  pos      = 'hbp!',  
  defaultOrientation = 'landscape') )
```

Other commands are available for mock listings

Mock Tables, *cont.*

R for Clinical
Trial
Reporting

Vanderbilt
Biostatistics

Software
Quality and
Validation
Quality and
Error Sources
What is Called
"Validation" and
What Should it
Be?
Example of a
Comprehensive
Analysis
Validation

High-Level
Tools for
Reproducible
Analysis and
Reporting

Background
Tools
Mock Tables
Statistical
Methods
Example

File `mock.dat` (actually one column):

```
*Baseline demographics      *Baseline hematology and clinical chemist
m Demographics              m Hematology
p Gestational age $<37$wc   RBC!$10^{12}$$/L
c Age                       c WBC!$10^{9}$$/L
p Female                   c Hgb!g/L
m Race                     c Hct!\%
p Caucasian                c Platelet count!$10^{9}$$/L
p Black                    m Clinical Chemistry
p Asian                   c ALT!U/L
p Hispanic                 c AST!U/L
p Native Am               c ALP!U/L
p Other                    c Bilirubin!$\mu$mol/L
```

Mock Tables, *cont.*

R for Clinical
Trial
Reporting

Vanderbilt
Biostatistics

Table: Baseline demographics, Age \leq 30d Group

	Placebo <i>N</i> = ##	0.01 mg/kg <i>N</i> = ##	0.1 mg/kg <i>N</i> = ##	1 mg/kg <i>N</i> = ##
Demographics				
Gestational age < 37w	##% (<i>n</i>)	##% (<i>n</i>)	##% (<i>n</i>)	##% (<i>n</i>)
Age	a b c	a b c	a b c	a b c
Female	##% (<i>n</i>)	##% (<i>n</i>)	##% (<i>n</i>)	##% (<i>n</i>)
Race				
Caucasion	##% (<i>n</i>)	##% (<i>n</i>)	##% (<i>n</i>)	##% (<i>n</i>)
Black	##% (<i>n</i>)	##% (<i>n</i>)	##% (<i>n</i>)	##% (<i>n</i>)
Asian	##% (<i>n</i>)	##% (<i>n</i>)	##% (<i>n</i>)	##% (<i>n</i>)
Hispanic	##% (<i>n</i>)	##% (<i>n</i>)	##% (<i>n</i>)	##% (<i>n</i>)
Native Am	##% (<i>n</i>)	##% (<i>n</i>)	##% (<i>n</i>)	##% (<i>n</i>)
Other	##% (<i>n</i>)	##% (<i>n</i>)	##% (<i>n</i>)	##% (<i>n</i>)

Software
Quality and
Validation
Quality and
Error Sources
What is Called
"Validation" and
What Should it
Be?
Example of a
Comprehensive
Analysis
Validation

High-Level
Tools for
Reproducible
Analysis and
Reporting

Background
Tools
Mock Tables
Statistical
Methods
Example

Mock Tables, *cont.*

R for Clinical
Trial
Reporting

Vanderbilt
Biostatistics

Software
Quality and
Validation

Quality and
Error Sources

What is Called
"Validation" and
What Should it
Be?

Example of a
Comprehensive
Analysis
Validation

High-Level
Tools for
Reproducible
Analysis and
Reporting

Background

Tools

Mock Tables

Statistical
Methods
Example

Table: Baseline hematology and clinical chemistry, Age ≤ 30 d Group

		Placebo	0.01 mg/kg	0.1 mg/kg	1 mg/kg
		N= ##	N= ##	N= ##	N= ##
Hematology					
RBC	$10^{12}/L$	a b c	a b c	a b c	a b c
WBC	$10^9/L$	a b c	a b c	a b c	a b c
Hgb	g/L	a b c	a b c	a b c	a b c
Hct	%	a b c	a b c	a b c	a b c
Platelet count	$10^9/L$	a b c	a b c	a b c	a b c
Clinical Chemistry					
ALT	U/L	a b c	a b c	a b c	a b c
AST	U/L	a b c	a b c	a b c	a b c
ALP	U/L	a b c	a b c	a b c	a b c
Bilirubin	$\mu\text{mol}/L$	a b c	a b c	a b c	a b c

Graphical Method for Interpreting Differences

R for Clinical
Trial
Reporting

Vanderbilt
Biostatistics

Software
Quality and
Validation

Quality and
Error Sources

What is Called
"Validation" and
What Should it
Be?

Example of a
Comprehensive
Analysis
Validation

High-Level
Tools for
Reproducible
Analysis and
Reporting

Background

Tools

Mock Tables

Statistical
Methods

Example

- Confidence limits have more information than 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

$$\frac{\bar{Y}_1 - \bar{Y}_2}{se} > z$$

$$\bar{Y}_1 - \bar{Y}_2 > z \times se$$

$$\text{Width of CL} = 2 \times z \times se$$

Example

R for Clinical
Trial
Reporting

Vanderbilt
Biostatistics

Software
Quality and
Validation

Quality and
Error Sources

What is Called
"Validation" and
What Should it
Be?

Example of a
Comprehensive
Analysis
Validation

High-Level
Tools for
Reproducible
Analysis and
Reporting

Background
Tools

Mock Tables
Statistical
Methods

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.