



Graphics for Clinical Trials

Frank E Harrell Jr

Department of Biostatistics
Vanderbilt University School of Medicine

ENAR 2015

Miami

2015-03-16



- Clinical trial reports have many standard components
 - accrual summary
 - patient flow/exclusions
 - baseline description
 - longitudinal analyses
 - adverse events
 - lab safety parameters (blood, ekg, etc.)
 - event timing/incidence
 - sequential monitoring of event probabilities
- Recognized need by reviewers to modernize clinical trial reports
- Tables do not lead to pattern perception



- Statisticians have made major progress in
 - reproducible research
 - statistical presentation graphics
- Graphics $>$ tables when > 2 numbers in the table
- The best graphics use features humans most accurately perceive: position along a common scale (relative length)
- Reproducible research requires scripting
- \Rightarrow need high level language to describe clinical trial report components



- Need for signposts on graphics
- Tables are secondary; can be in an appendix and hyperlinked
- Emphasize confidence intervals for *differences*
- Show entire distributions when possible
- Favor quantiles over moments
- Percentages are inherently confusing; always replaced with proportions and ratios



- R, \LaTeX , knitr, Hmisc, lattice, ggplot2
- New R package greport
 - utility functions
 - high-level report component functions
 - unified handling of figure generation, captions
 - new graphical implementations



New Graphical Elements

- Extended box plots
- Special dot charts for stratified proportions
- Half-violin plots (vertical density plots)
- Half-confidence intervals
 - Centered at midpoint of two estimates
 - Length = $\frac{1}{2}$ length of CL
 - \Rightarrow touches two estimates \Leftrightarrow difference not significant at $\alpha = 0.05$
- Pop-up tooltips for explanation of graphic or underlying table
- Spike thermometers as signposts



Assumptions

- Treatment variable has the same name in every data frame
 - Define as `tx.var` option to `setgreportOption()`
- Subject randomized \Leftrightarrow randomization date not NA
- There is an **official** number of enrolled and randomized subjects (called `denom`)
- Almost every variable has a `label`
- Continuous variables have `units`
- Whenever a dataset being analyzed has > 1 record/subject, an `id` variable must appear in the formula specifying the analysis
- Some of the graphics only handle two treatments for now



Features

- Hyperlinks: graphics link to tables which link back to graphics
- Automatic composition of figure captions
- Captions for graphs have table numbers in appendix
- Captions for tables have figure numbers in report body
- Analysis specification: analysis variables on left-hand-side and stratification and id variables on right-hand-side of formula



`Merge` : merge multiple `data.frames` or `data.tables`, providing a table describing what happened

`setgreportOption` : set uniform graphics symbols, color, line types, line thickness for treatments; `tx.var`, denominators, whether hyperlinks used, figure positioning, sub-directory name to hold generated \LaTeX code and pdf graphics

`getgreportOption` : retrieve setup

`upFirst` : capitalization of first letter of each *appropriate* word in a phrase



Functions, *continued*

accrualReport : subject accrual, taking into account regions, countries, sites

exReport : exclusion and improper randomization report

dReport : descriptive statistics for baseline and longitudinal data

eReport : event report

survReport : time-to-event report

nriskReport : number-at-risk report (declining denominators for longitudinal data)

Future addition : sequential monitoring of event incidence/stopping boundaries



Example Output of Merge

	Vars	Obs	Unique IDs	IDs in #1	IDs not in #1
demog	6	1080	1080	NA	NA
enroll	6	1080	1080	1080	0
rand	2	604	604	604	0
randie	2	579	579	579	0
ccta1	6	690	690	690	0
. . . .					
Merged	83	1080	1080	1080	0

Number of unique IDs in any data frame : 1080

Number of unique IDs in all data frames: 0



Graphics for
Clinical Trials

Background

Tools

Assumptions

Features

Functions

This work used only free software

L^AT_EX





A New Model for Reproducible Clinical Trial Statistical Reporting

Frank Harrell

Statisticians and statistical programmers spend a great deal of time analyzing data and producing reports for clinical trials, both for final trial reports and for interim reports for data monitoring committees. Point and Click interfaces and copy-and-paste are now believed to be bad models for reproducible research. Instead, there are advantages to developing a high-level language for producing common elements of reports related to accrual, exclusions, descriptive statistics, adverse events, time to event, and longitudinal data.

It is well appreciated in the statistical and graphics design communities that graphics are much better than tables for conveying numeric information. There are thus advantages for having statistical reports for clinical trials that are almost completely graphical. For those reviewers of clinical trial reports who insist on seeing tables, and for those who occasionally like to have tables to see "exact" figures for certain data elements, supporting tables can be placed in an appendix. These tables are hyperlinked to the main graphics. Small tables can also pop-up when one hovers the mouse over a graphic. These two approaches are facilitated by .



features of Adobe Acrobat Reader. Reviewers who prefer printed reports can print the appendix in order to have a complete document.

In this talk I will describe a new R package called greport (“graphical report”) that marries R, the R `Hmisc`, `ggplot2`, and `lattice` packages, `knitr`, and \LaTeX to produce reproducible clinical trial reports with a minimum of coding. `greport` composes all figure captions and makes heavy use of analysis file annotations such as variable labels and units of measurement. Some new graphical elements are introduced such as special dot charts that replace tables, extended box plots, split violin plots for longitudinal continuous variables, half confidence intervals for differences, new charts for representing patient flow, and pop-up tooltips. Supporting tables are hyperlinked to graphics, and the graphics are hyperlinked back from the tables. Figure captions contain supporting table numbers, and tables contain figure numbers. Two example reports will be shown.

For more information see <http://biostat.mc.vanderbilt.edu/Greport>.