
Graphical Representation of DataFax Reports Using R- Project Statistical Software

Zekai Otles
D. Gaston, P. Lenon, and X. Li
Frontier Science & Technology Research Foundation
Madison, WI-USA
otles@fstrf-wi.org



Objectives

- Display or tabulate DataFax reports in a comprehensible fashion to determine patterns, trends, or discrepancies as a part of better clinical trials data management.
- Eliminate the needs for additional resources (SAS, MS-Excel) or transferring data.
- Generate tools to display data using R.



Background

- Gulati V. and Green K. (DFUG-2001); Using DataFax Reports to Monitor and Improve Site Performance.
- R is an open source statistical computation and graphics software.
<http://www.r-project.org/>



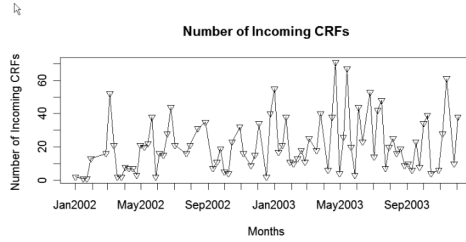
Introduction

- The data used here are fictional.
- The DataFax report results and R-graphs are not produced from the same dataset due to the confidentiality.
- The tools used in this presentation were developed using only R.

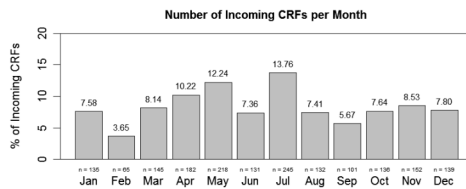


Ex: Weekly incoming CRFs (DF_WFcrfperwk)

2002:43	10 *
2003:13	6 *
2003:16	1
2003:17	4
2003:50	1
2004:05	8 *
2004:10	2
2004:51	2
2005:13	1
2005:14	1
2005:22	7 *
2005:23	12 *
Total	55 pages in 12 weeks
Mean	5 pages per week

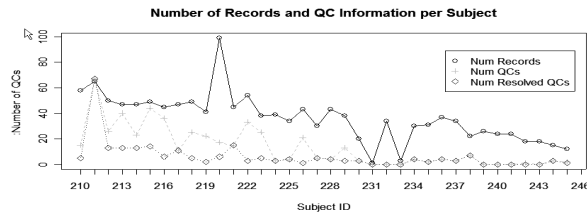


- Time plots are easily comprehensible
- easy to modify
- Bar plots demonstrate the patterns



Ex: Number of QCs per subject (DF_PTqcs)

PATIENT ID#	TOTAL RECORDS	TOTAL QCS	RATE/100 REC	RESOLVED QCS	AGE (DAYS) OF UNRESOLVED QCS					
					#	%	<=10	11-20	21-30	31-60
10001	4	1	25	0	0	0				1
10002	5	0	0	0	0	0				
10111	1	2	200	0	0	0				2
10121	1	0	0	0	0	0				
20001	6	0	0	0	0	0				
30001	6	0	0	0	0	0				
TOTAL	23	3	13	0	0	0	0	0	0	3

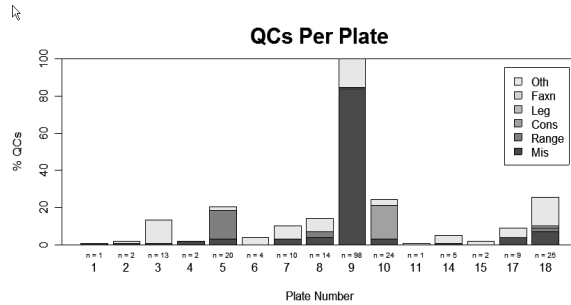


- Easy to monitor subject's data retrieval process and QCs
- Snap shot of the clinical trial data



Ex: Number of QCs per plate (DF_qcsbyfield)

PLATE FIELD DESCRIPTION	TOTAL	MISS	RANG	CONS	LEGI	FAXN	OTHR
1 7. Problem_Description	1:	0	1	0	0	0	0
1 9. Problem_Description	1:	0	1	0	0	0	0
1 21. Test2nd	1:	0	1	0	0	0	0
1 23. Test3rd	1:	0	1	0	0	0	0
2 9. Entry Date	1:	0	0	1	0	0	0
2 11. Date of Birth	2:	0	0	2	0	0	0



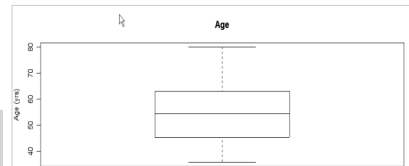
- Existing QCs can either be displayed for each plate as shown above or as fields within a plate
- Inform PIs on error-prone plate or fields on a specific plate
- Data Monitor can train sites on expected trouble plate/s or field/s on a plate



Ex: Simple statistics on demographic value age

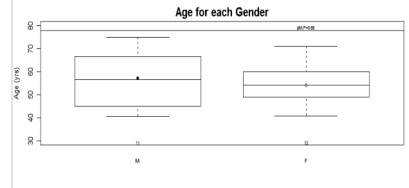
DF_Stats For Plate 1 DataFax Test Study Jan 05,2006 14:38

8. PINIT: Patient Initials	6	100.00%
Text Found	3	3
Text Length		
9. AGE: Age	6	100.00%
Legal Values		
Mean	52.83	
Standard Deviation	7.19	
Range	45-61	
10. SEX: Sex	3	50.00%
1 = male		
2 = female	3	50.00%



TRF	TOTAL	PTS	Mean	STD	DEV	Median	Q1	Q3	P25	P75
N	11		52.83	7.19	52.83	45.00	49.50	56.00	48.50	75.00
F	13		54.15	10.48	54.00	49.00	60.00	49.80	71.00	

- More statistical tools available
- Displays data graphically or in tabular format



Ex: How to use this tool from dataFax reporting system

The screenshot shows the dataFax reporting system interface. On the left, there are menu options: 'File', 'Preferences', 'History'. Below these are 'Report Type' (DataFax Generic, Study Specific, History List) and 'crfPerWk: Display Number of Incoming CRFs'. There are 'Options', 'Run Report', and 'Explain' buttons. The main window displays a report titled 'Number of Incoming CRF' with a line graph showing data from June 2002 to January 2003. Below the graph is a bar chart titled 'Number of Incoming CRFs per 1' with values: 7.50, 8.85, 8.14, 10.02, 12.24, 7.20, 13.76, 7. The DFUG logo is in the bottom right corner.

-Displays how to run this system

R software development

```
#####
# Function crfBarPlot
#####

"crfBarPlot" <- function(shMatrix,titleOfPlot=NULL)
{
  barValsLoc=30 #it locates the bar values location maxVal/barValsLoc
  scaleCex=0.5
  # Label bars with their values
  sumVal<-sum(shMatrix[,1])
  perVals<- 100*(shMatrix[,1]/sumVal)
  maxPerVal<-max(perVals)
  barVals<-array(1:length(perVals))

  #identify the ymax for the ylimits
  if (maxPerVal<=91){
    #set the ticks for next 10th value
    yLimMax<-10*ceiling(maxPerVal/10)
  }else{yLimMax<-100}

  yOffset=1.5
  #give space to write last value
  deltaY<-yOffset+1.5*scaleCex
  if ((yLimMax -maxPerVal)<deltaY) {yLimMax<-yLimMax+deltaY}

  bp<-barplot(perVals,ylim=c(0,yLimMax),ylab="% of Incoming CRFs")

  for(i in 1: length(barVals)) {
    barVals[i]<-sprintf("%2.2f",perVals[i])
  }
  box()
  text(x=bp, y=yOffset+perVals, cex=(1.5*scaleCex), labels=barVals)
  #text(titleOfPlot,side=3,line =3, at=titleLoc,cex=3*scaleCex, font = 2,adj=0.5)
  #text(titleOfPlot,side=3,line =1, cex=2*scaleCex, font = 2,adj=0.5)
  #text(paste("n = ",shMatrix[,1], sep=""),side=1,line=0,at=bp,cex=1.0*scaleCex)

  > library("dfTools")
  > load("datafax.Rdata")
  > crfPerWk(sh)
```



Summary

- R-software used to develop graphical tools.
- Our R-based tools can help better manage clinical trials.
- This tool eliminates data transfer among systems.
- Tools can be run in the study specific report.



Remarks

- This tool can help
 - Study PIs
 - Monitor accrual rates
 - Identify or correct where QCs are (site or CRF based) and what causes them
 - Creates snap shots and quality of their clinical data
 - Provide simple statistics in a graphical or table format
 - Data Manager
 - Helps determine the work load
 - Quickly identifies sites needing further training or complicated plates requiring extra detailed attention
 - Clinical Research Associates
 - Helps to decrease QC rates by paying attention to the specific plate/s or field/s

