The Pharmaceutical Industry's Polyamorous Relationship with R

Luke Fostvedt

Pfizer, Inc.
Global Product Development

Enterprise Applications of the R Language (EARL) Conference
San Francisco, California, USA

June 7, 2017

Introduction

R is making much headway in the Pharmaceutical Industry and is commonly used during the development of investigational drugs/treatments.

- Identifying mutations/gene for a new compound to target
- Quality control when manufacturing the compound
- Designing Clinical trials
- Modeling and Analysis of the Data from our clinical trials
- and many other applications

Some Traditional software used in drug development

Much of the software variety is due to people wanting to use the softwares with which they were trained.

- ▶ SAS, East, PASS: Primarily used by Statisticians
- ▶ NONMEM, PsN, Phoenix, ADAPT: Primarily used by pharmacologists
- ▶ Minitab, JMP, Design Expert, Matlab: Used by various researchers
- ▶ Other out-of-the-box softwares: Subject specific

Drug Development

Pre-clinical work:

- Gene Associations/mutations
- in vitro [and in vivo] work
- Animal Trials

Clinical Trials:

- Phase 1 Studies
  - First in Human
  - Dose Finding
  - Safety
  - Pharmacology Studies
  - Bioequivalence Studies
  - Dose Finding
- Phase 2 Studies
  - Dose Finding
  - Efficacy/Safety
- Phase 3 Studies
  - Randomized Controlled Trial

New Drug Application:

- Dosage Recommendations
- Side Effects/Risks
- and many more components

Pre-Clinical

- Bioinformatics
    - Bioconductor packages are widely used
    - Some people use SAS
- Animal Studies
    - Mostly GUI based software
    - Quantitative Systems Pharmacology (ADAPT, Berkeley-Madonna, Phoenix, etc)

Drug Formulation

If SAS/JMP/Design Expert/Minitab can not handle some predictive modeling easily or make ideal plots, R is commonly used as an alternative because of its flexibility.

- ► Tolerance interval on specification setting:
  - ► library(tolerance)
- ► Visualization:
  - ► library(ggplot2)
- ► Linear mixed effect models:
  - ► library(lme4)
  - ► library(nlme)
- ► Bayesian methods:
  - ► library(rjags)
- ► Data manipulation:
  - ► library(reshape)
  - ► library(reshape2)
  - ► library(plyr)

Model Based Meta Analyses

Model-based meta-analysis can help teams understand how a disease or compound is working enabling improved decision making regarding dosing, adverse events, efficacy, whether to continue developing a specific compound.

- R Base, plyr for data manipulation, post-processing (alternate software: SAS, SPlus)
- pcnetmeta for star plot
- ggplot2 for visualization (alternate software: SPlus, Spotfire)
- knitr and pander for reporting (alternate software: LateX, Word)
- metafor for traditional meta-analysis, bias checks and forest plots (an alternate software: SAS)
- stats (glm function) , lme4 (glmer function) for modeling (alternate Software: NONMEM, SPlus)
- boot for model diagnostics (alternate software: SAS)
- R2OpenBUGS or rjags for Bayesian modeling (alternate software: SAS, STAN)

Clinical Trial Design

Traditionally SAS has been used to design and power clinical studies. However other out-of-the-box Software such as PASS (NCSS) or EAST (Cytel) are commonly used.

- First in Human dose finding studies
  - library(BOIN)
  - library(dfcomb)
- Powering Bioequivalence studies
  - library(PowerTOST)
  - library(Power2Stage)
- Group Sequential Designs
  - library(gsDesign)
- Model Based Adaptive Optimal Design
  - library(MBAOD)
- User-written simulations for trial designs

First In Human Studies

In oncology rule-based trial designs are commonly used for dose-finding
clinical trials

- Traditional 3+3 design remains the prevailing method for conducting
  phase I cancer clinical trials
- Model-based adaptive designs are also possibilities

MD Anderson Cancer Center provides (for free)

- R packages and R code for designing and analyzing first-in-human
  clinical trials
    - Bayesian Optimal Interval Designs: library(BOIN)
    - Adaptive Dose Modification (ADM)
    - Bayesian Dose-finding Designs for Drug Combination: library(dfcomb)
    - many more
- Other home-made software
    - http://odin.mdacc.tmc.edu/

Survival Analyses

Survival Analyses can be done with SAS, R, JMP, Minitab, SPSS, etc.

- the `survival` library provides functions for most common survival analyses
  - Kaplan-Meier Analyses
  - Cox Proportional Hazard Models
  - Various survival distributions
  - model diagnostics
  - etc

Pharmacology Work

Traditionally NONMEM, PsN (PearlspeaksNonmem), Phoenix (and other Certara products), Berkeley-Madonna, ADAPT, and many others have been the primary software used for Pharmacokinetic and Pharmacodynamic (PKPD) work. Since PKPD modeling typically involves solving ordinary differential equations, much effort has been done to either convert NONMEM code to R or vice versa.

- ▶ Noncompartmental Pharmacokinetic (PK) Analysis
  - ▶ At least 5 packages on CRAN
  - ▶ Many companies have their own internal code/shiny apps
- ▶ Population PKPD Modeling
  - ▶ library(mrgsolve)
  - ▶ library(nlmixr)
  - ▶ library(rspeaksnonmem)
  - ▶ etc

Patent Applications

I recently had to provide simulations to the lawyers for a patent application. They wanted me to simulate several different dosing regimens and provide summary tables.

- ▶ Using R I created several datasets corresponding to different dosing regimens
- ▶ I used the NONMEM model to simulate the exposures with PsN from my datasets
- ▶ I used R for post-processing
- ▶ ggplot2 for the visualizations

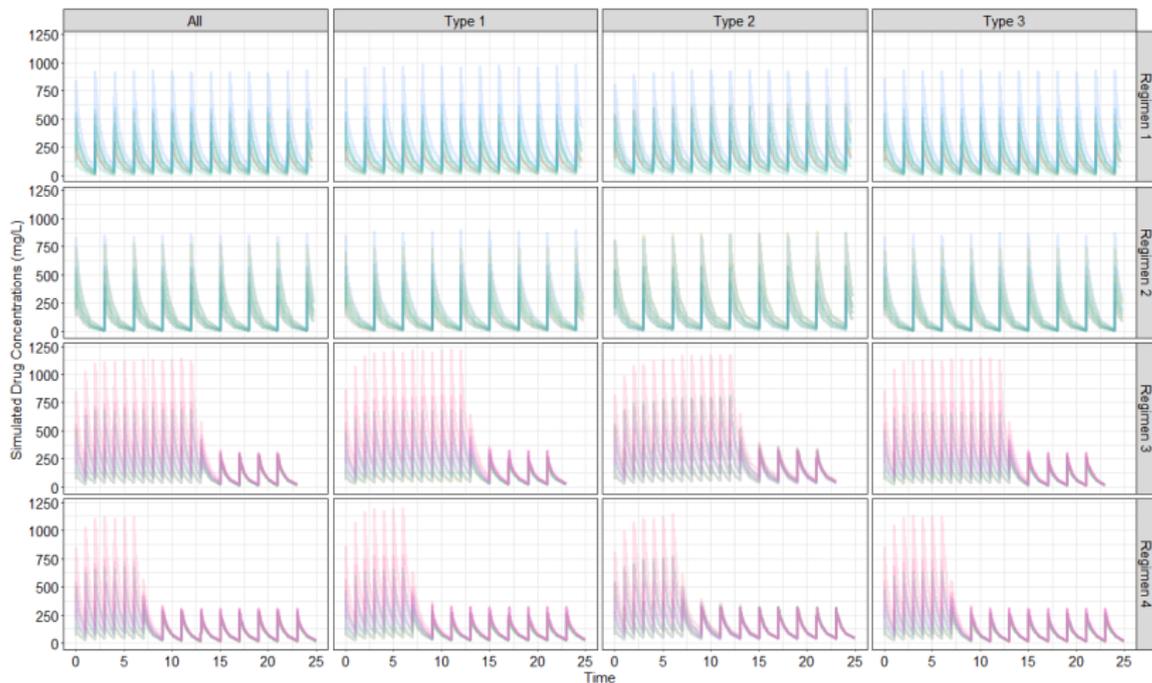But I could have done everything with R :)

Patent Applications

Figure 1:Simualted exposures from different dose regimens

Shiny Apps

Shiny Apps are being developed across all product development parts of the pharmaceutical industry. Both to help those who are not programmers (eg MDs, lab scientists, etc) and to simplify/standardize processes. Some examples are

- ▶ Gene Associations
- ▶ Data review/cleaning
- ▶ Non-Compartmental Analysis of Pharmacokinetics
- ▶ Others

Shiny apps are being developed that can compete with Tibco's Spotfire.

Example 1: Shiny App for Data Review/Cleaning

Throughout the life of a clinical trial it is important that the data we are collecting is clean and accurate.

- Reviewing the data "dumps" at scheduled intervals (eg monthly) enables us to identify sites that are not collecting or reporting the data as defined in the Protocol
- Since these dataset can get quite large, using Excel (what non-R programmers commonly use) to review the data can be overwhelming.
- Shiny apps (or other R code) can help to streamline the review process to catch common errors
- Particularly visualizations of the data can be helpful
- Using the reactive objects and brushing functions for graphics helps identify "unusual" values

Shiny App for Electrocardiogram (ECG) Data Review
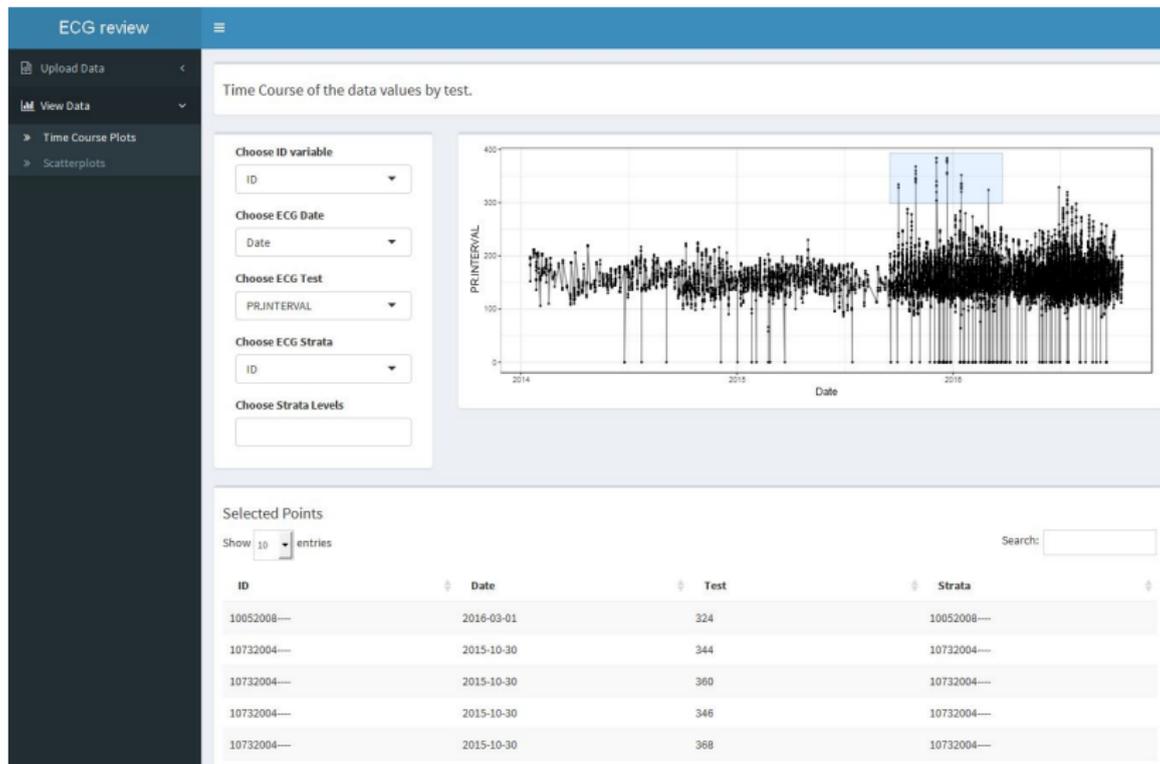

Figure 2:Some of the measurements are unusually large

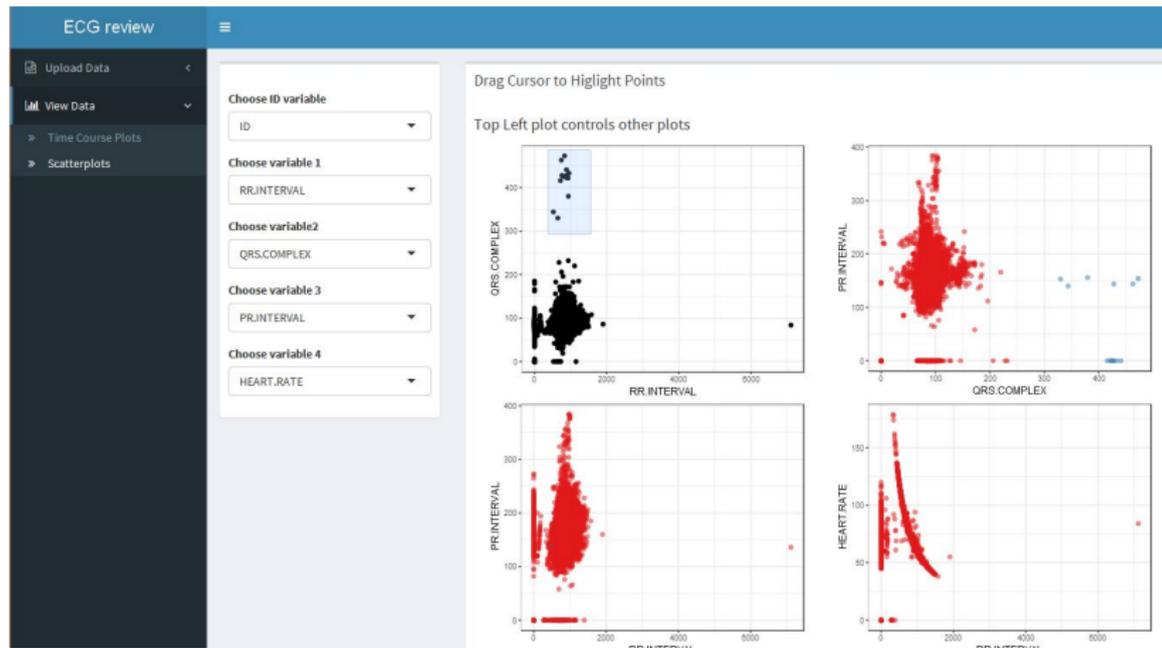# Shiny App for Electrocardiogram (ECG) Data Review

Figure 3:Some of the pairwise associations are unusual

Example 2: Using R for a Bioequivalence Study

Everytime a new drug formulation is created, a bioequivalence study is conducted to ensure the drug exposure of the new formulation is not significantly different from the old formulation.

- ► How consistent is the production of the new formulation?
  - ► Quality control analyses can easily be performed with R
- ► What Pharmacokinetic exposure data do we have from the previous formulation regarding intrasubject variability and intersubject variability?
  - ► several R packages available for the NCA PK estimates
  - ► several packages for nonlinear mixed effects modeling

Example 2: Using R for a Bioequivalence Study

- How can we design and power the study?
  - packages such as 'powerTOST, Power2Stage, and others
  - can also be used to analyze the result
- Create dataset from the data collected from the study
  - base R functions
  - data manipulation packages (eg dplyr, reshape, tidyr, etc)
- Create tables and figures for the Clinical Study Report (CSR)
  - ggplot2 and other standard data manipulation packages
- Draft final report
  - R markdown

Example 3: PKPD Report Generation

I need to do an exposure-response analysis looking at Progression Free Survival (PFS).

- Which studies have PK and PD information available for designing my dataset for my analysis?
- What do I know about the exposure from the Population PK analysis model?
- What software will I use to draft my report?

Available Studies

Identify relevant studies that have collected pregression free survival (PFS).

- Using Excel list of variables is provided to the programming group for dataset creation.
- SAS is used to create the dataset
- Review each iteration using R and whichever packages I find helpful (ggplot2, dplyr, survival, etc.)
- Cross reference the dataset with the Tables and Listings (produced with legacy SAS code) from the Clinical Study Reports (CSRs)
- A final dataset is created using SAS

Analysis and Modeling

In many cases, I need to simulate patient exposure data.

- ▶ Using the Population PK model I simulate the relevant exposure metrics (Maximum concentration, cumulative exposure. etc)
- ▶ This simulation is commonly conducted using NONMEM
- ▶ post-processing is completed using R
- ▶ Once I have the exposure data, I add it to my dataset for my analysis using the `survival` package in R
- ▶ All tables and figures are created in R using packages such as `ggplot2`.
- ▶ The final report is then generated using LaTeX which compiles all the tables and figures I created into the final document

Conclusions

- ▶ R is quickly gaining traction across all areas and stages of drug development
- ▶ R is flexible enough to be used almost entirely thoughout the entire drug development process
- ▶ This presentation was created with R markdown