Using R in a Regulatory Environment: some FDA perspectives

Paul Schuette
Scientific Computing Coordinator
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics
Paul.Schuette@fda.hhs.gov

Disclaimer

This presentation reflects the views of the author and should not be construed to represent the FDA's views or policies.

Statistical Software Clarifying Statement

"FDA does not require use of any specific software for statistical analyses, and statistical software is not explicitly discussed in Title 21 of the Code of Federal Regulations [e.g., in 21CFR part 11]. However, the software package(s) used for statistical analyses should be fully documented in the submission, including version and build identification.

As noted in the FDA guidance, *E9 Statistical Principles for Clinical Trials*, 'The computer software used for data management and statistical analysis should be reliable, and documentation of appropriate software testing procedures should be available.' Sponsors are encouraged to consult with FDA review teams and especially with FDA statisticians regarding the choice and suitability of statistical software packages at an early stage in the product development process. "

https://www.fda.gov/downloads/forindustry/datastandards/studydatastandards/ucm587506.pdf

R for Regulatory Review

How is R used for regulatory review work?

- Reviewers may opt to perform their analyses using R rather than commercial packages.
- R is used for graphics and data visualization.
- Simulations in general.
- Bayesian Methods
  - JAGS
  - Stan
- Complex, Innovative Clinical Designs (PDUFA VI)

Some R packages for Biostatistics

- survival, Therneau
- Hmisc, Harrell *et al*
- DoseFinding, Bornkamp, Pinheiro, and Bretz
- gsDesign, Anderson
- Beanz, Wang *et al*
- ORCI, Sun

IDE RStudio is used extensively at FDA.

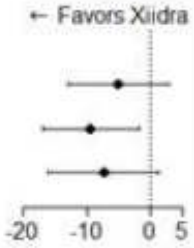
Product Label

https://www.accessdata.fda.gov/drugsatfda_docs/label/2016/208073s000lbl.pdf

Figure 1: Mean Change (SD) from Baseline and Treatment Difference (Xiidra – Vehicle) in Eye Dryness Score in 12-Week Studies in Patients with Dry Eye Disease

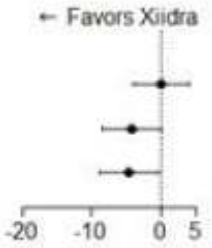
Study 1

Visit	Vehicle (N = 58)	Xiidra (N = 58)	Difference[1] (95% CI)
Baseline	51.8 (23.55)	51.6 (24.69)	
Day 14	-3.9 (25.46)	-8.9 (21.72)	-5.1 (-13.1, 3.0)
Day 42	-7.9 (19.60)	-17.3 (24.96)	-9.4 (-17.0, -1.9)
Day 84	-7.2 (25.29)	-14.4 (25.36)	-7.3 (-16.1, 1.4)



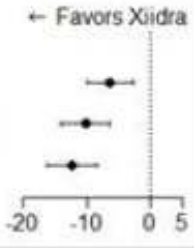
Study 2

Visit	Vehicle (N = 295)	Xiidra (N = 293)	Difference[1] (95% CI)
Baseline	41.6 (29.69)	40.2 (28.64)	
Day 14	-7.5 (29.01)	-6.7 (27.36)	0.1 (-3.9, 4.1)
Day 42	-9.1 (30.03)	-12.6 (30.71)	-4.2 (-8.5, 0.0)
Day 84	-11.2 (28.78)	-15.2 (31.48)	-4.7 (-8.9, -0.4)



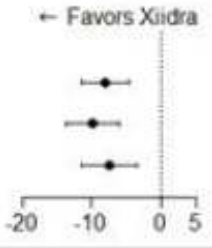
Study 3

Visit	Vehicle (N = 360)	Xiidra (N = 358)	Difference[1] (95% CI)
Baseline	69.2 (16.76)	69.7 (16.95)	
Day 14	-13.1 (24.04)	-19.7 (26.49)	-6.4 (-10.0, -2.8)
Day 42	-18.2 (26.51)	-28.3 (27.69)	-10.0 (-13.8, -6.1)
Day 84	-22.8 (28.60)	-35.3 (28.40)	-12.3 (-16.4, -8.3)



Study 4

Visit	Vehicle (N = 356)	Xiidra (N = 355)	Difference[1] (95% CI)
Baseline	69.0 (17.08)	68.3 (16.88)	
Day 14	-14.9 (22.35)	-22.7 (25.41)	-8.0 (-11.4, -4.5)
Day 42	-23.7 (25.98)	-33.0 (27.46)	-9.6 (-13.4, -5.8)
Day 84	-30.5 (28.03)	-37.7 (28.91)	-7.5 (-11.6, -3.5)



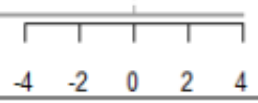
[1] Based on ANCOVA model adjusted for baseline value in Study 1, and ANCOVA model adjusted for baseline value and randomization stratification factors in Studies 2-4 All randomized and treated patients were included in the analysis and missing data were imputed using last-available data. In Study 1, one Xiidra treated subject who did not have a baseline value was excluded from analysis.

Another Product Label

R Graphic. Drug for the reduction of elevated intraocular pressure (IOP) in patients with open-angle glaucoma or ocular hypertension
https://www.accessdata.fda.gov/drugsatfda_docs/label/2017/208254lbl.pdf

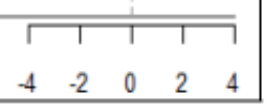
Study 304: Subjects with Baseline IOP < 25 mmHg

Visit	Rhopressa (N=186)	Timolol (N=187)	Difference (95% CI) Rhopressa - Timolol
Baseline			
8am	22.4	22.4	
10am	21.1	21.3	
4pm	20.7	20.7	
Change From Baseline			
Day 15			
8am	-4.7	-4.9	0.2 (-0.4, 0.8)
10am	-4.5	-4.5	0.0 (-0.5, 0.5)
4pm	-4.4	-3.8	-0.6 (-1.1, -0.1)
Day 43			
8am	-4.6	-4.8	0.3 (-0.3, 0.8)
10am	-4.3	-4.3	-0.1 (-0.6, 0.5)
4pm	-4.1	-4.0	-0.1 (-0.6, 0.4)
Day 90			
8am	-4.5	-5.2	0.6 (0.0, 1.2)
10am	-4.1	-4.5	0.4 (-0.2, 0.9)
4pm	-3.9	-3.9	0.0 (-0.6, 0.5)



Study 304: Subjects with Baseline IOP >= 25 and < 30 mmHg

Visit	Rhopressa (N=120)	Timolol (N=130)	Difference (95% CI) Rhopressa - Timolol
Baseline			
8am	26.3	26.0	
10am	25.2	24.9	
4pm	24.5	24.0	
Change From Baseline			
Day 15			
8am	-4.7	-5.9	1.2 (0.3, 2.0)
10am	-5.0	-5.6	0.6 (-0.2, 1.5)
4pm	-4.3	-4.9	0.6 (-0.2, 1.3)
Day 43			
8am	-4.3	-6.2	1.9 (1.0, 2.8)
10am	-4.7	-5.8	1.1 (0.2, 1.9)
4pm	-4.3	-4.4	0.2 (-0.6, 1.0)
Day 90			
8am	-4.5	-6.1	1.6 (0.6, 2.5)
10am	-4.1	-5.9	1.8 (0.9, 2.7)
4pm	-3.9	-5.0	1.1 (0.2, 1.9)



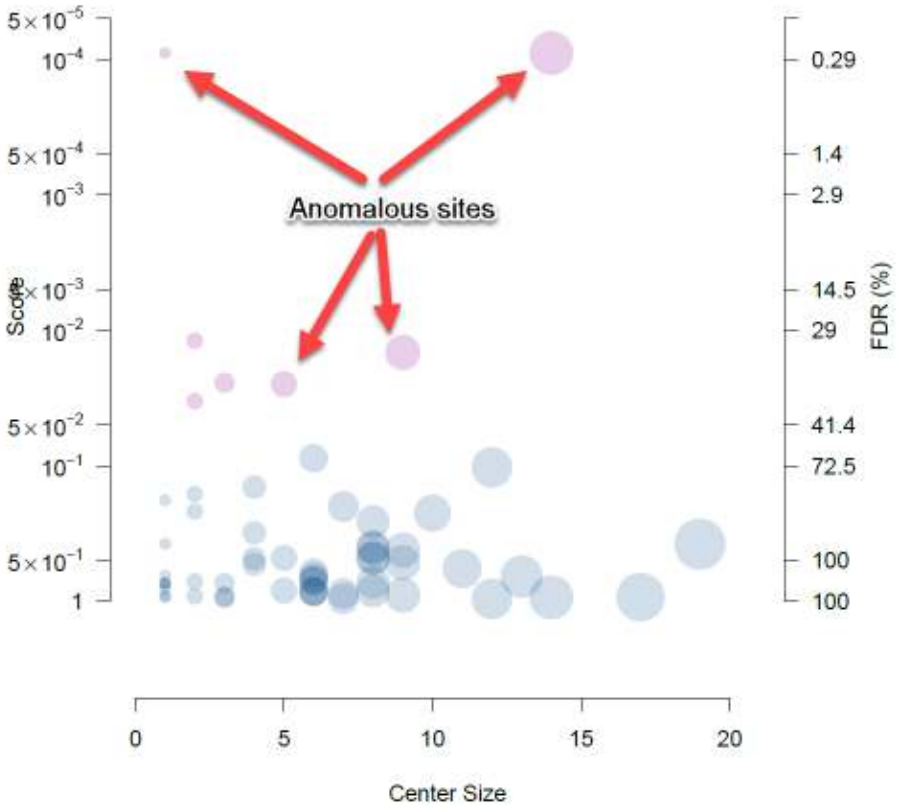
This table was produced based on the observed data from all randomized subjects who did not have major protocol violations. The treatment differences and two-sided CIs for comparing Rhopressa QD vs Timolol BID 0.5% were based on Analysis of Covariance (ANCOVA) adjusted for baseline IOP.

Data Anomaly Detection

Use open source software to detect potential data problems

1. DABERS: **D**ata **A**nomalies in **B**io**E**quivalence **R** **S**hiny app. Used for PK/PD profiles.
2. Cooperative Research and Development Agreement (CRADA) with CluePoints for detecting anomalous clinical trial sites.

Example of CRADA software output



R Shiny Apps

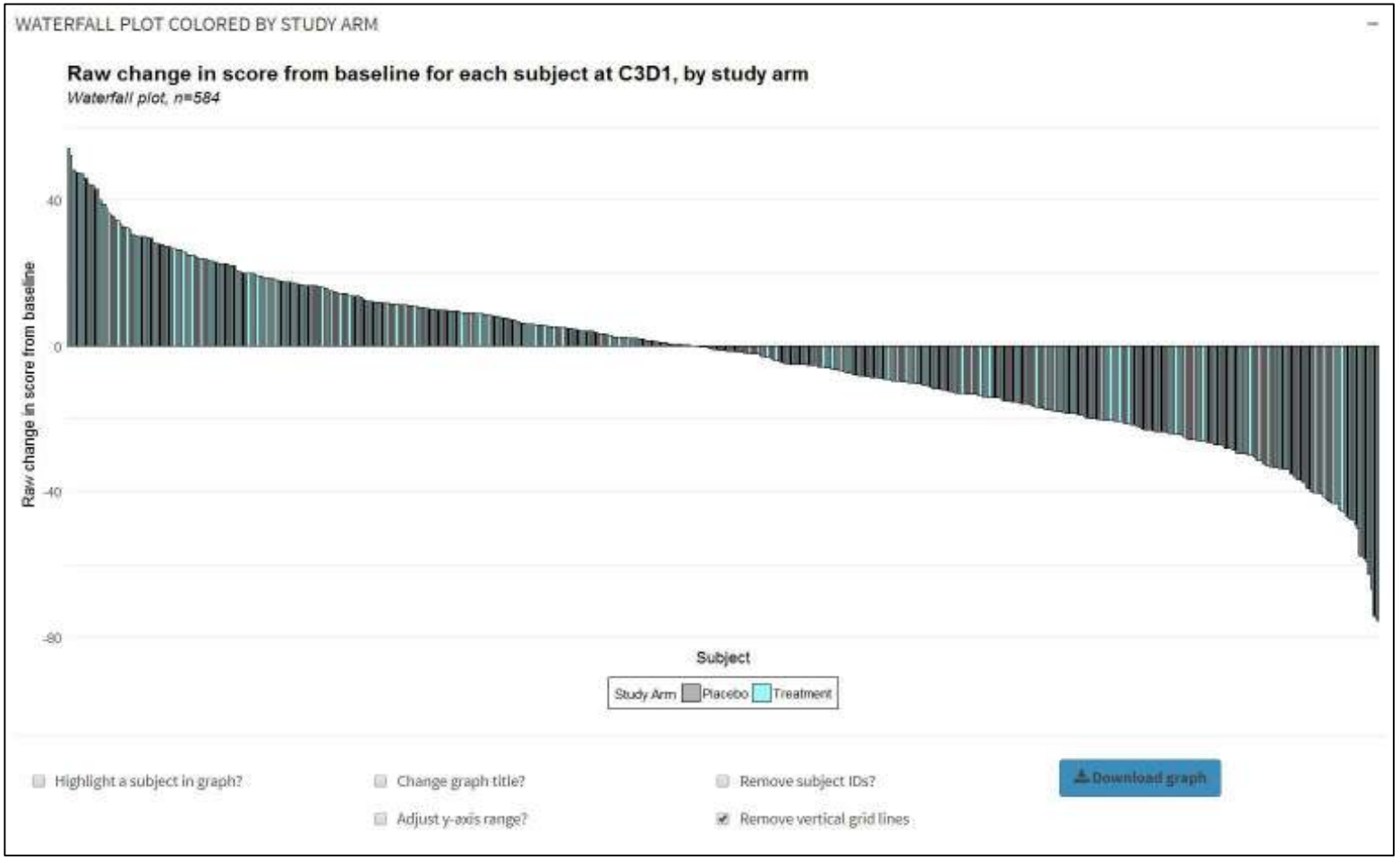
Internal to FDA

- Waterfall Plot
- Hepatotoxicity
- Demographics
- PRO
- DABERS

External to FDA (openFDA)

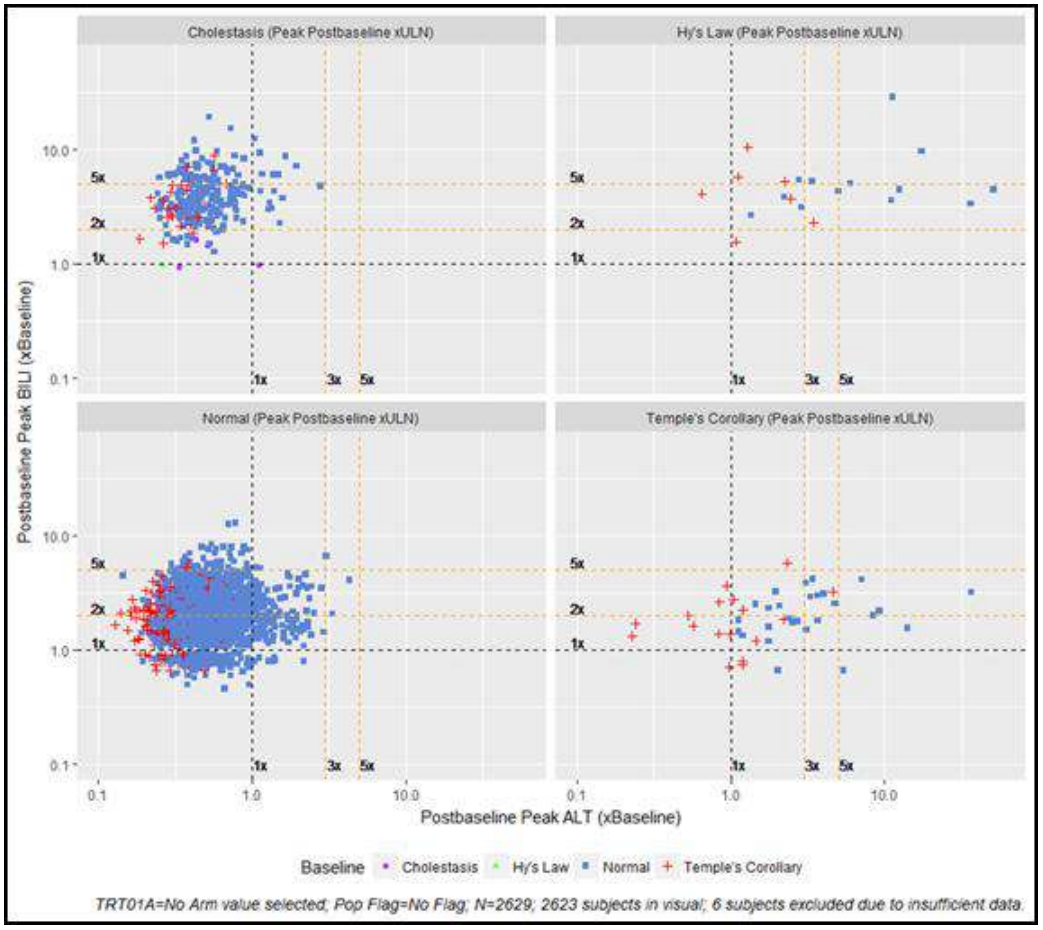
- LRT app for Adverse Event analyses

Waterfall Plot



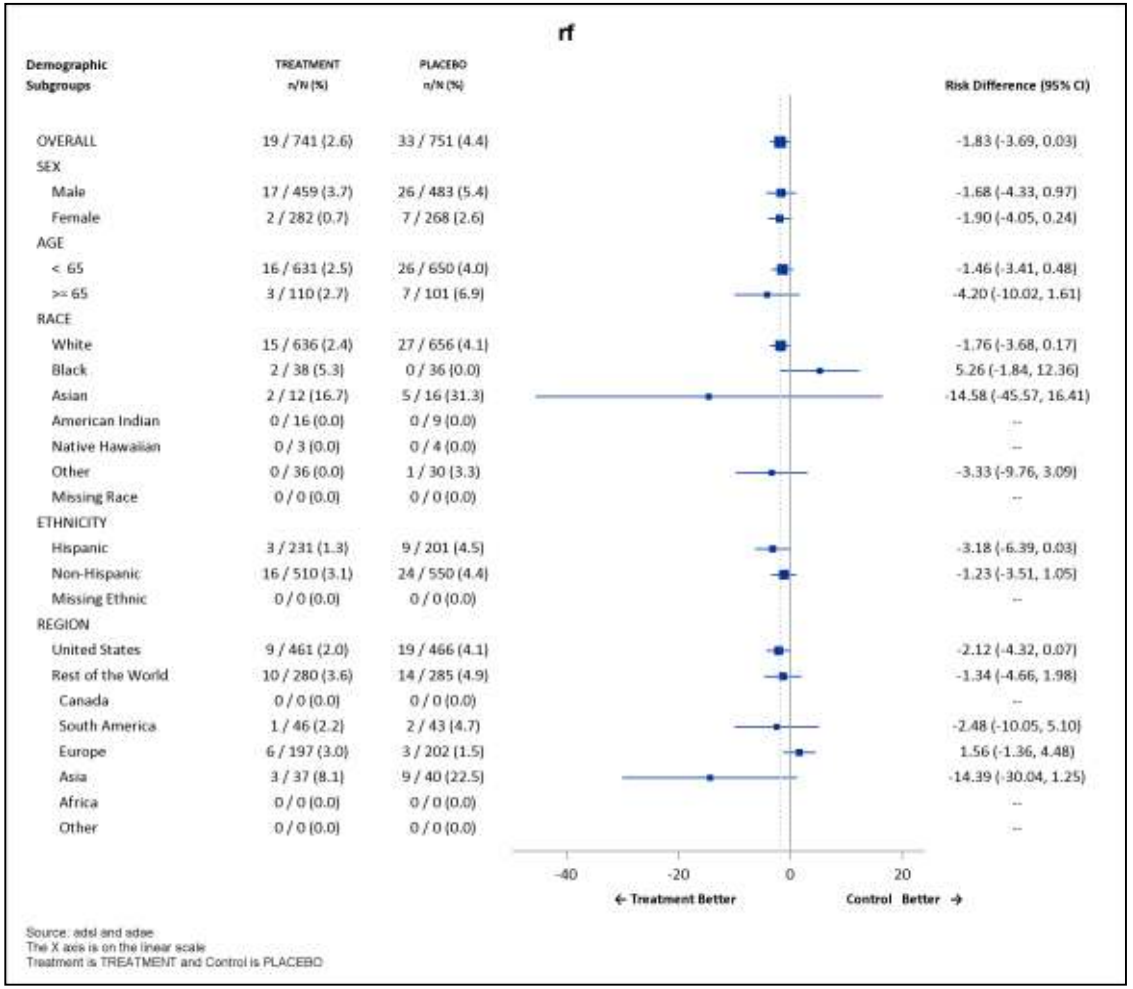
Hepatotoxicity

The Hepatotoxicity tool bolsters analysis of Drug Induced Liver Injury (DILI) through a composite visualization that includes both pre-treatment and on-treatment prevalence of ALT and BILI in terms of Hy's Law candidate laboratory Upper Limit Normal (ULN) thresholds as well as the magnitude of these elevations normalized by respective baseline test results. This analysis is particularly useful for studies in which subjects have elevated liver enzyme test results at baseline (e.g., subjects with Chronic Hepatitis C).



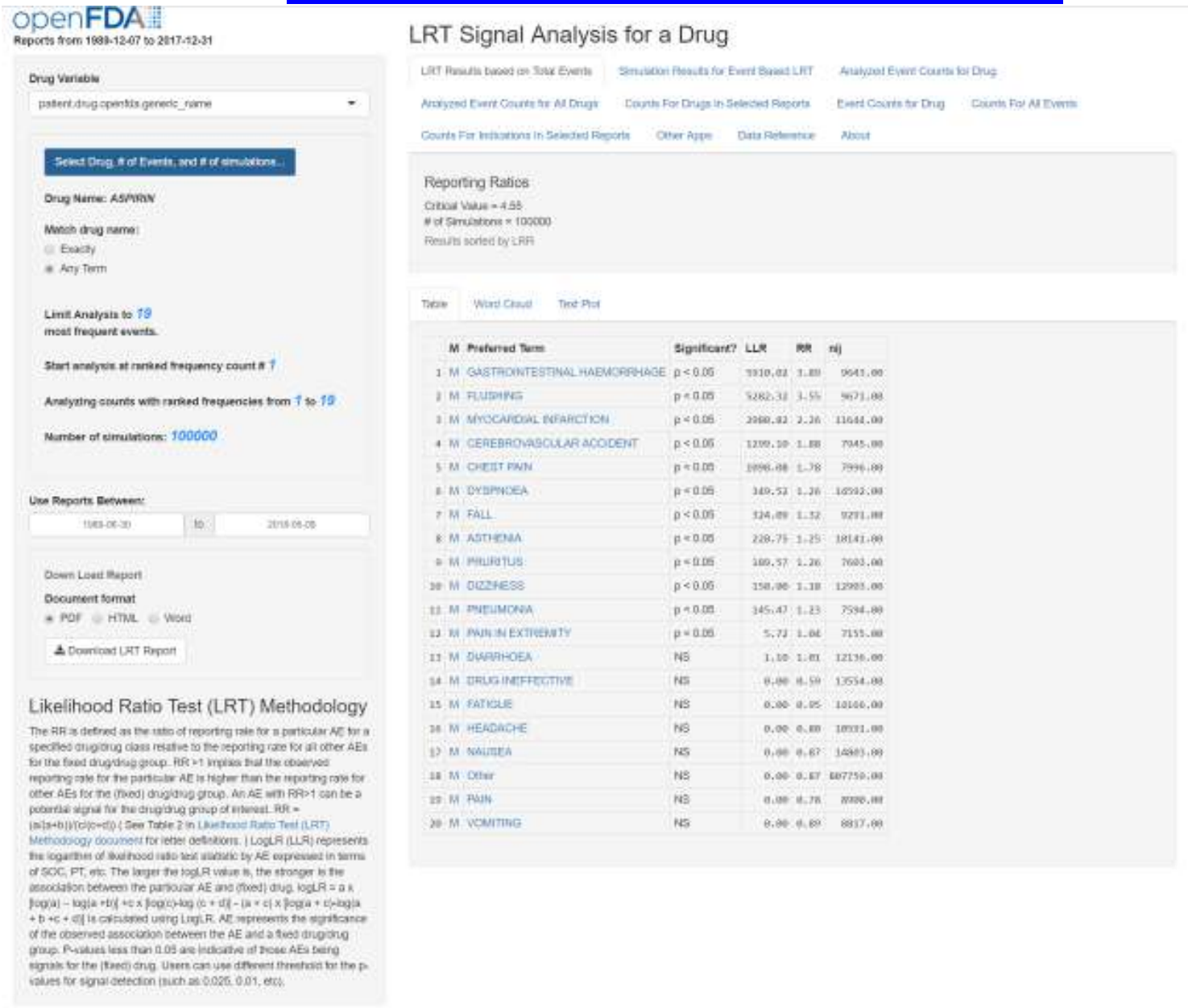
Demographic Tool

The Demographic Tool provides targeted descriptive statistics and safety endpoint analysis for demographic subgroups, including age, sex, race, and ethnicity. The tool has a simple user interface that dynamically walks end-users through the process of executing the analysis. The example deals with a safety endpoint analysis.



FAERS data, OpenFDA

https://openfda.shinyapps.io/LRTest/



openFDA

Reports from 1989-12-07 to 2017-12-31

Drug Variable

patient.drug.openfda.generic_name

Select Drug, # of Events, and # of simulations...

Drug Name: ASPIRIN

Match drug name:

Exactly

Any Term

Limit Analysis to 19 most frequent events.

Start analysis at ranked frequency count # 1

Analyzing counts with ranked frequencies from 1 to 19

Number of simulations: 100000

Use Reports Between:

1989-06-30 to 2016-06-06

Down Load Report

Document format

PDF HTML Word

Download LRT Report

Likelihood Ratio Test (LRT) Methodology

The RR is defined as the ratio of reporting rate for a particular AE for a specified drug/drug class relative to the reporting rate for all other AEs for the fixed drug/drug group. RR >1 implies that the observed reporting rate for the particular AE is higher than the reporting rate for other AEs for the (fixed) drug/drug group. An AE with RR>1 can be a potential signal for the drug/drug group of interest. RR = (a/(a+b))/(c/(c+d)) (See Table 2 in Likelihood Ratio Test (LRT) Methodology document for letter definitions.) LogLR (LLR) represents the logarithm of likelihood ratio test statistic by AE expressed in terms of SOC, PT, etc. The larger the logLR value is, the stronger is the association between the particular AE and (fixed) drug. logLR = a x [log(a) – log(a +b)] +c x [log(c)-log (c + d)] – (a + c) x [log(a + c)-log(a + b +c + d)] is calculated using LogLR. AE represents the significance of the observed association between the AE and a fixed drug/drug group. P-values less than 0.05 are indicative of those AEs being signals for the (fixed) drug. Users can use different threshold for the p-values for signal detection (such as 0.025, 0.01, etc).

LRT Signal Analysis for a Drug

LRT Results based on Total Events | Simulation Results for Event Based LRT | Analyzed Event Counts for Drug | Analyzed Event Counts for All Drugs | Counts For Drugs in Selected Reports | Event Counts for Drug | Counts For All Events | Counts For Indications in Selected Reports | Other Apps | Data Reference | About

Reporting Ratios

Critical Value = 4.55

# of Simulations = 100000

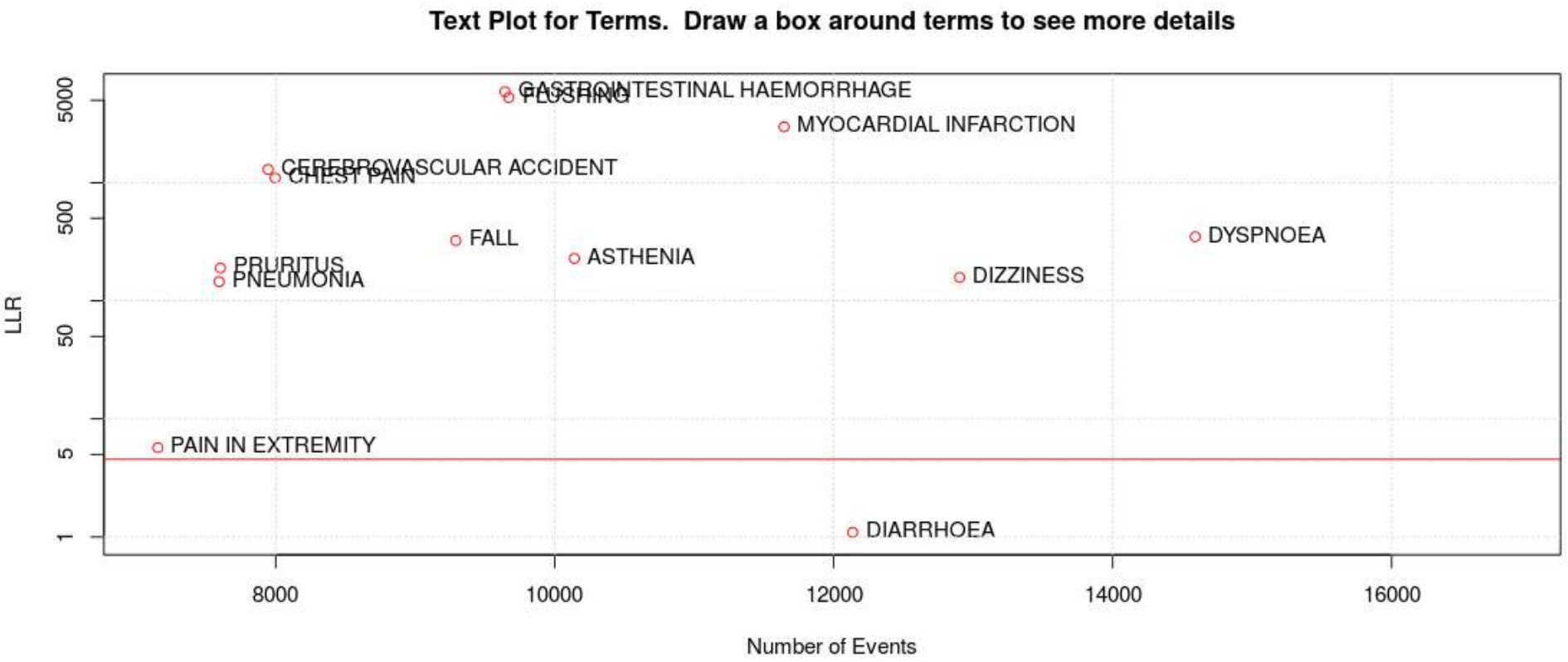
Results sorted by LRR

Table | Word Cloud | Text Plot

	M	Preferred Term	Significant?	LLR	RR	nij
1	M	GASTROINTESTINAL HAEMORRHAGE	p < 0.05	5910.02	3.89	9643.00
2	M	FLUSHING	p < 0.05	5282.32	3.55	9071.00
3	M	MYOCARDIAL INFARCTION	p < 0.05	2980.82	2.26	11644.00
4	M	CEREBROVASCULAR ACCIDENT	p < 0.05	1200.10	1.88	7045.00
5	M	CHEST PAIN	p < 0.05	1096.88	1.78	7996.00
6	M	DYSPNOEA	p < 0.05	349.52	1.26	16593.00
7	M	FALL	p < 0.05	324.89	1.32	9291.00
8	M	ASTHENIA	p < 0.05	228.75	1.25	18141.00
9	M	PRURITUS	p < 0.05	189.57	1.26	7603.00
10	M	DIZZINESS	p < 0.05	158.86	1.18	12903.00
11	M	PNEUMONIA	p < 0.05	145.47	1.23	7594.00
12	M	PAIN IN EXTREMITY	p < 0.05	5.72	1.06	7155.00
13	M	DIARRHOEA	NS	1.10	1.01	12136.00
14	M	DRUG INEFFECTIVE	NS	0.00	0.59	13554.00
15	M	FATIGUE	NS	0.00	0.05	10160.00
16	M	HEADACHE	NS	0.00	0.00	10931.00
17	M	NAUSEA	NS	0.00	0.67	14983.00
18	M	Other	NS	0.00	0.67	807750.00
19	M	PAIN	NS	0.00	0.76	8980.00
20	M	VOMITING	NS	0.00	0.69	8817.00

Text Plot from LRT app,

Drug: aspirin



Birthdate Problem

Birthdate! | Basic Birth Date Problem | With Initials | With last 4 SSN

Basic Birth Date Problem. What is the probability that at least two subjects in a group share the same date of birth (month, day and year)?

Age Range

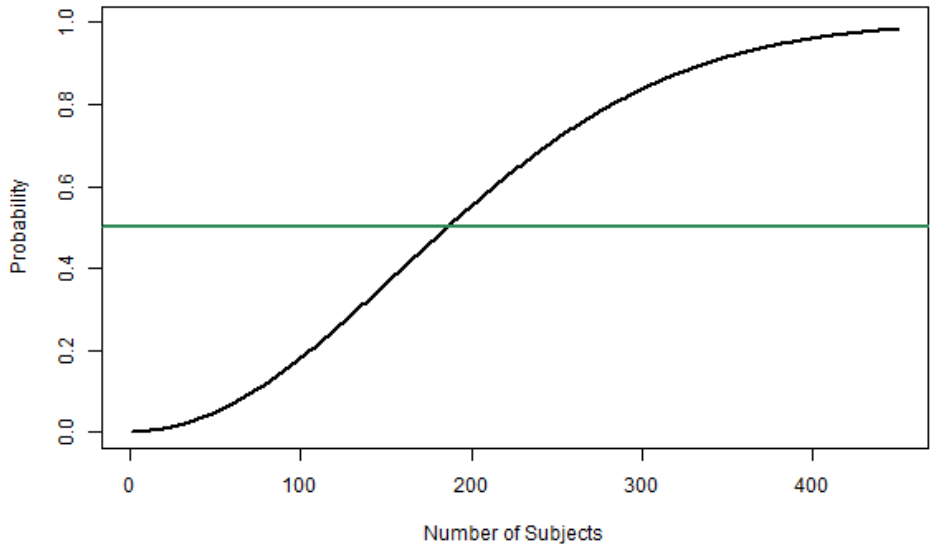
0 18 85 100

0 10 20 30 40 50 60 70 80 90 100

Probability Level

0 0.5 1

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

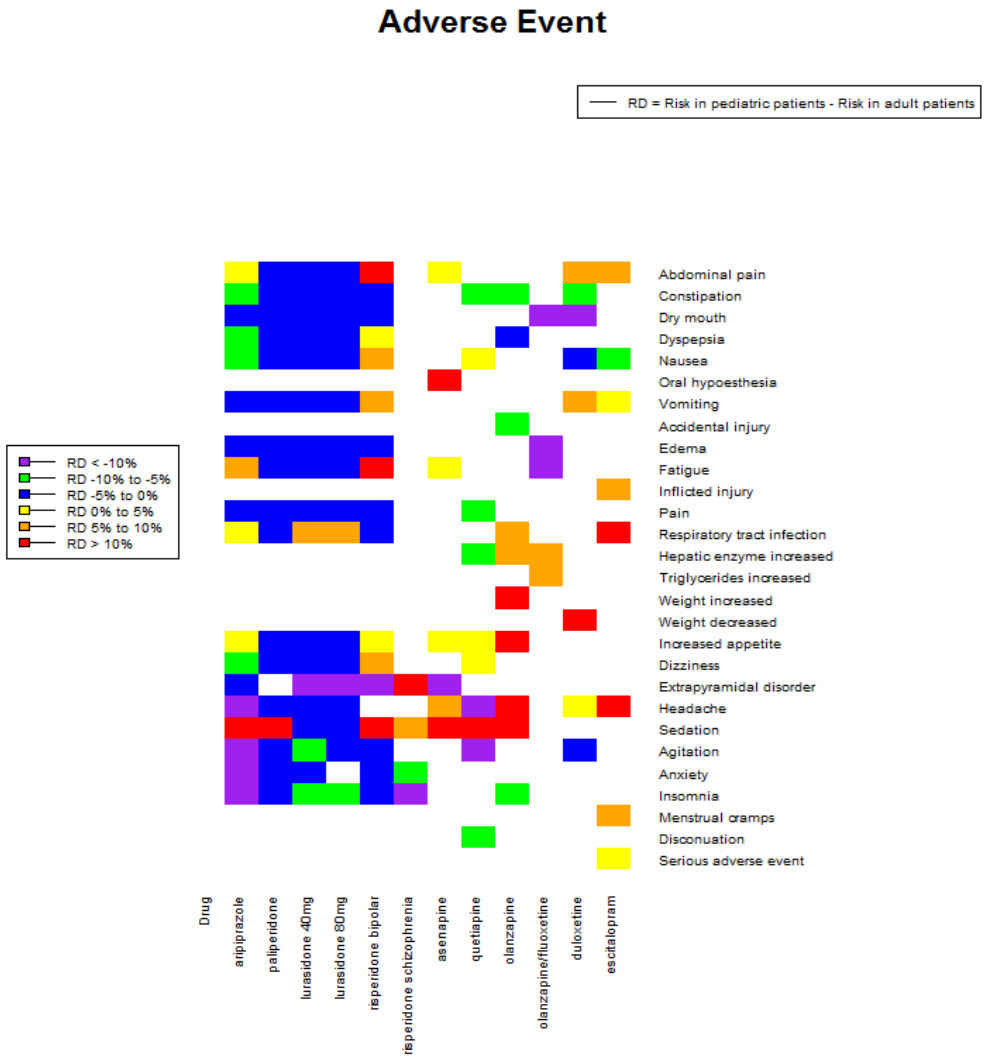


For probability level 0.5 , the required number of subjects is N= 186

R for Research

- Data Mining and Machine Learning (also with Python)
- Simulations
- Evaluation of methodology
- Oak Ridge Institute for Science and Education (ORISE) Internships
- Broad Agency Agreements (BAA)
- Cooperative Research and Development Agreements (CRADA)
- PhUSE, DIA, and ASA working groups

Research, Pediatric vs Adult ADRs



Concluding Observations

- Open source tools such as R offer cost effective ways for FDA to carry out its public health mission, and to enhance communications with the public, health care providers and regulated industry.
- R is widely used in academe, and is the first choice for many recent graduates.
- Managing packages and dependencies can be challenging.
- Interactive tools such as R Shiny can enhance users' experience and understanding.
- We still need subject matter experts to help frame questions and draw appropriate conclusions.

Acknowledgements

- Yan Wang
- Jimmy Wong
- Daniel Choi
- Austin Taylor
- Xiaomei Liu


FDA
U.S. FOOD & DRUG
ADMINISTRATION