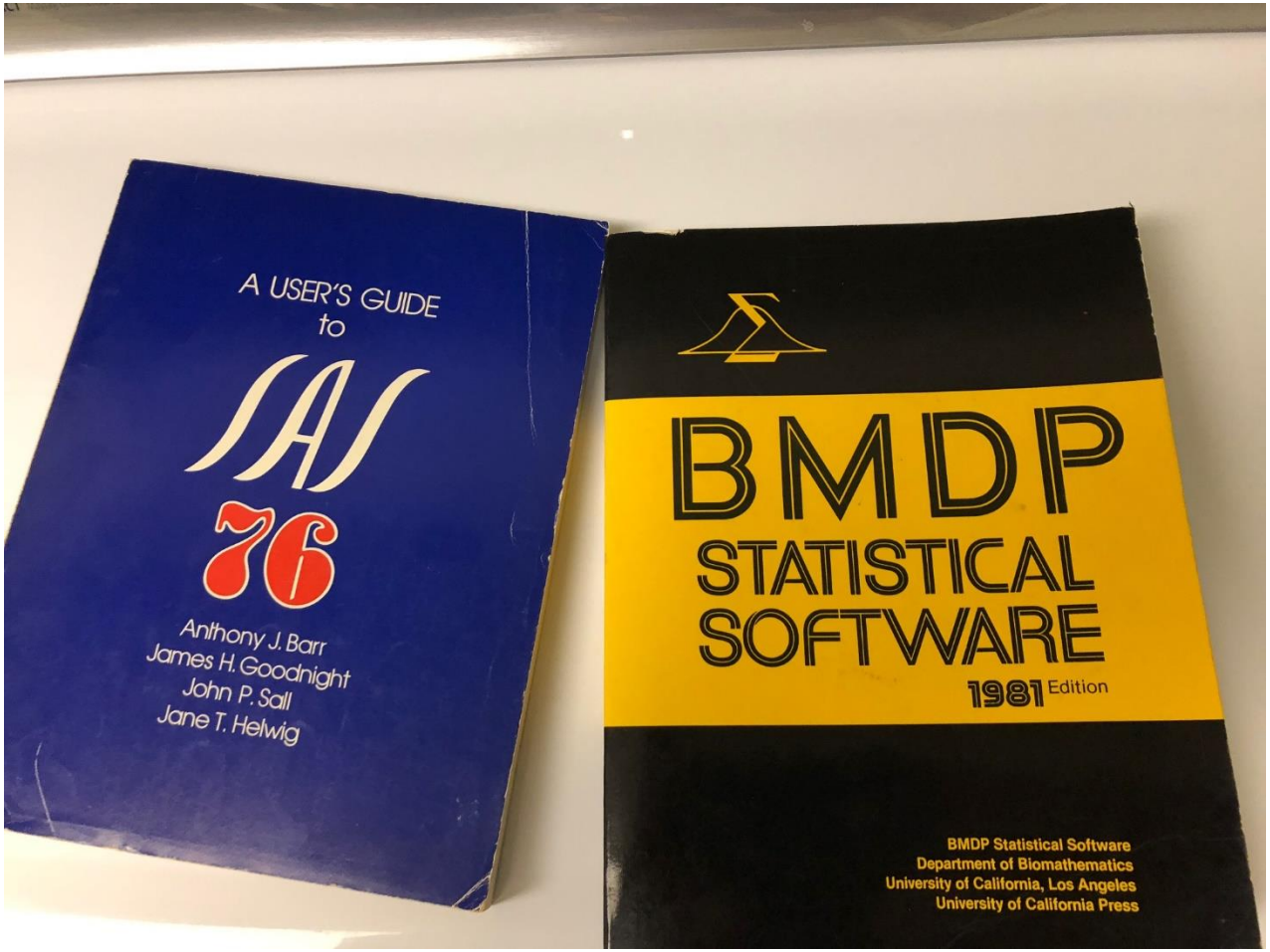
Infrastructure for World-Wide Clinical Trials: The BeiGene Case Study

Alan Hopkins PhD
VP, Global Biometrics
BeiGene Ltd

PharmaSUG China 2018
August 31, 2018

Early Days of Stat Computing



Historical Background: The Old Days (Circa 1982)

- Large Pharma used SAS for statistical analysis and so did the US FDA
- Biotech startups may not have an IBM mainframe.
- My first pharma job was to create tools for statistical analysis on a Unix machine.
  - BMDP statistical Fortran library to create customized BMDP-style software for study reports
  - P-Stat software did the work of the SAS data step on Unix
  - Used open-source S software for graphics
- Data "capture" was with paper forms double-entered into a generic SQL database
- Unix tools were used to manage the statistical programming environment

BeiGene Biometrics: Dec 2016

- Multiple studies underway with 3 oncology drugs
- Activities outsourced to CROs
- US: 2 Statisticians, 3 programmers
- China: 0 statisticians, 0 programmers
- One shared compute server with limited bandwidth and storage

Goal: Build an organization where the clinical data pathway is managed by BeiGene Biometrics.

Key Components to Manage the Clinical Data Pathway

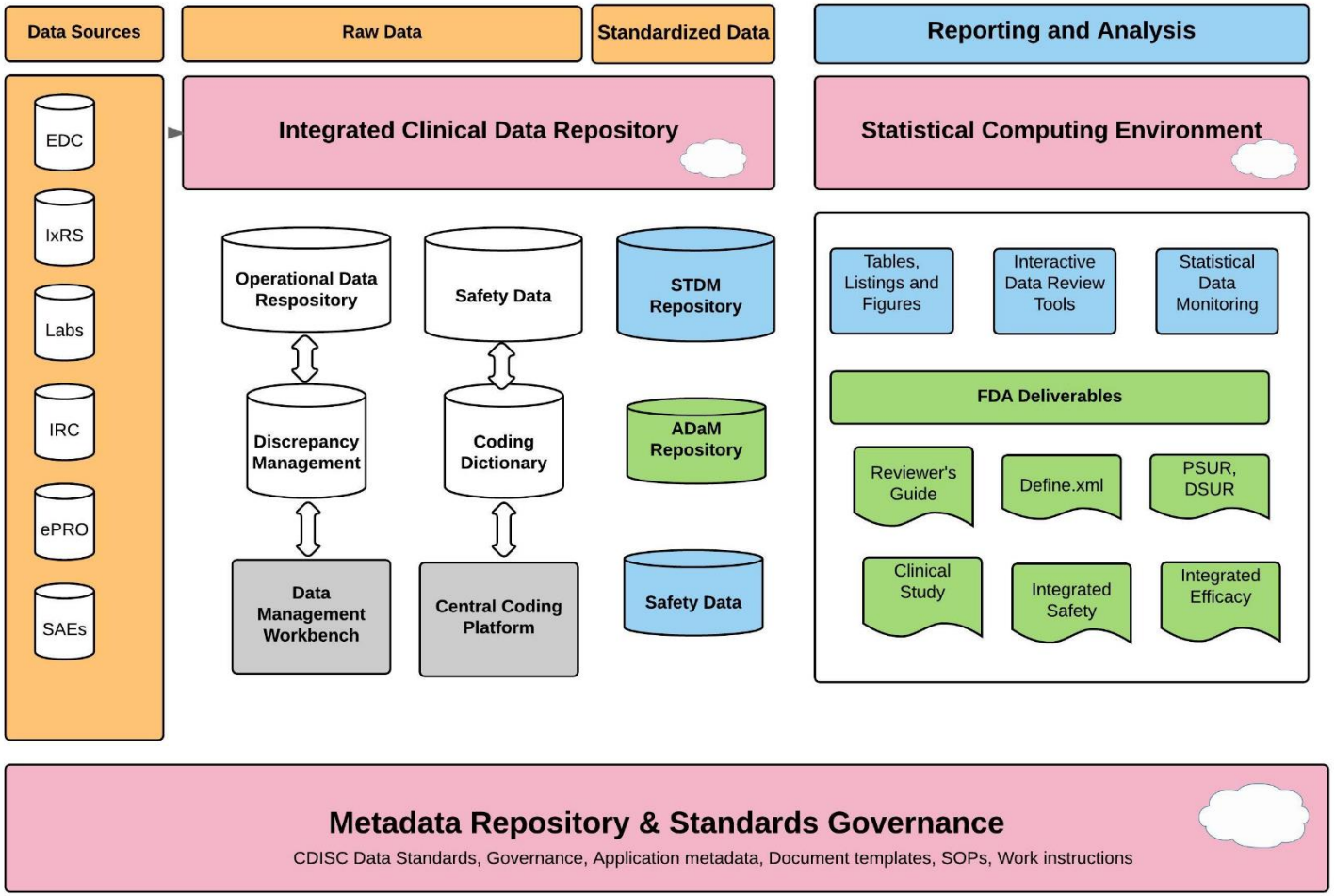
1. **Statistical Computing Environment (SCE)[1]**
   - Software employs version control, audit trails, traceability, dependency management, impact analysis, role-based security, workflow and tasks to create an environment for clinical reporting and analysis.
2. **Integrated Clinical Data Repository (CDR)**
   - Centralize, secure, store, share and deliver operational and study data.
   - Sophisticated versioning combined with traceability between the received and transformed data along with all related assets ensures that the lineage, integrity and provenance of data are always available and reproducible.
3. **MetaData Repository (MDR) framework: Data Standards**
   - Collaborate with stakeholders on metadata specification and drive reuse in study build by connecting to EDC systems. Refine your standards using MDR Governance.
   - Accelerate EDC builds, data transformations and statistical programming
4. **DART – Data Analysis and Reporting tools**
   - Software tools including SAS, Spotfire, R for data review/visualization

[1]Hopkins, Duke & Dubman (2010) Statistical Computing Environments and the Practice of Statistics in the Biopharmaceutical Industry. Therapeutics Innovation & Regulatory Science.

Statistical Computing Environment Features and Benefits

Feature	Benefit
Process automation	Create repeatable, enforceable, predictable processes
Version control and accountability	History of all documents including statistical analysis source code, data files, logs, and output, and establishes chain of custody
Dependency management	Impact analysis and update process management (
Document repository	Store documents associated with programs, data, and analysis
Metadata management	Insure consistency in use and meaning of data and analysis; manage all metadata used in processing of information, work flow, and change control
Submission metadata publishing	Create reports such as define.xml data documentation, automated tables of contents, footnotes, and titles
Program status flags	Allow enforcement of business rules concerning validation
Scalability	Ability to support small or large organizations

Clinical Data Lifecycle Management: Where's the Data?



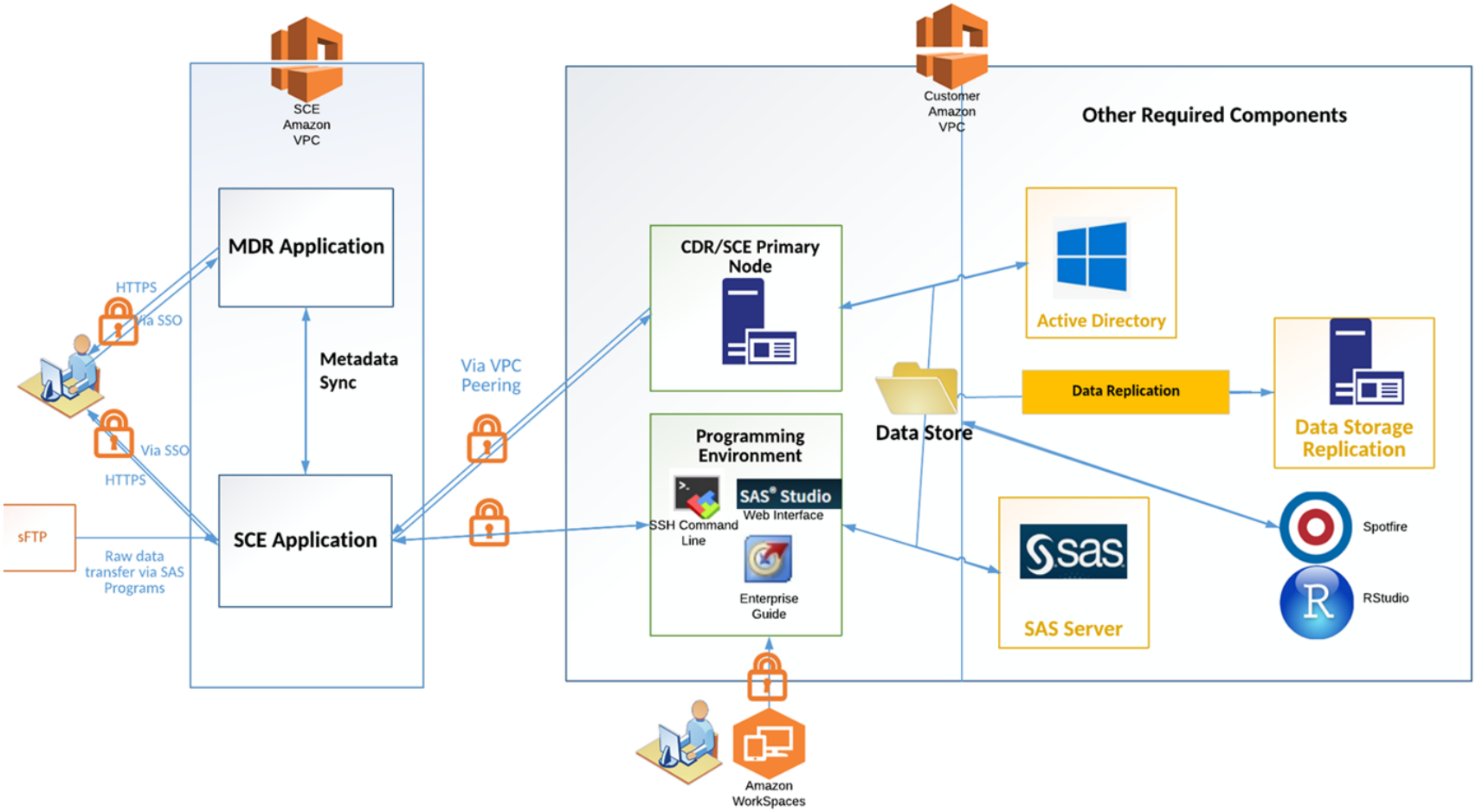
Amazon Web Services Platform (AWS): Why Cloud Infrastructure?

- High-Performance and Scalable
  - Wide Performance Range: Small Web Servers to High Compute (I/O) Servers
  - On Demand Scaling – Increase or decrease capacity within minutes
  - Storage capabilities is easily extensible
- Reliable
  - Built-in redundancy through Amazon's infrastructure
  - Globally Accessible
  - Disaster Recovery capability built in
- Flexible
  - Supports wide range of OS, Application Platforms, Databases, Storage, Networking etc.
  - Works in conjunction with other software vendor cloud platforms
- Secure
  - Secure technology platform built and managed by Amazon
- Cost Effective
  - Pay-as you go pricing
  - Fewer FTEs to deploy and support
  - No need for building, maintaining or upgrading own infrastructure

Implementation Scope

1. **Implement secure and globally accessible compute infrastructure on Amazon Web Services (AWS)**
   - Compute servers for statistical computing environment
   - Implement prerequisite components (Active Directory, Identity Mgmt., VPN, standard laptops)
   - Controlled virtual desktop environment in AWS co-located with the compute servers
2. **Implement validated data analysis and reporting tools (DART)**
   - SAS Platform : SAS Compute, SAS Studio, SAS EG, SAS Office Analytics
   - R : R Server Studio Server and R
   - Spotfire Data Review and visualization tool
3. **Implement a validated secure file exchange platform**
   - Secure sFTP repository/staging area for CRO/vendor data to be retrieved and pushed to SCE environment
4. **Framework for MDR/CDR/SCE platform**
   - Any middleware or connectors required to integrate with the MDR/CDR/SCE platform and/or data warehouse

High Level Hardware Architecture

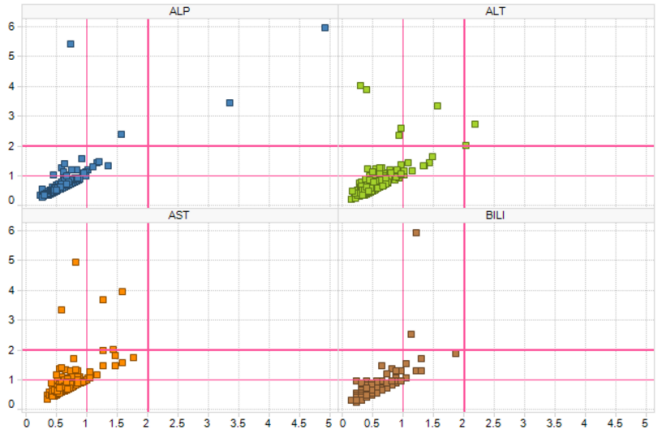


Spotfire in Clinical Development

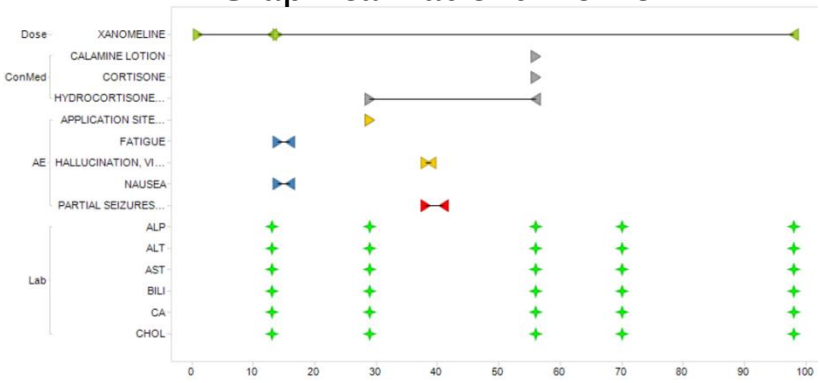
Clinical Development

- In stream Clinical Data Review
- Trial Operations/Risk Based Monitoring
- Drug Safety and Pharmacovigilance
- Data Management Review & Queries

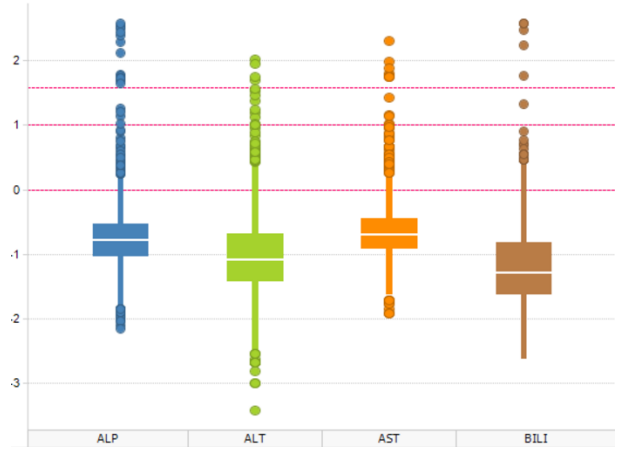
Hy's Law – Liver Lab Outliers



Graphical Patient Profile



Lab xULN



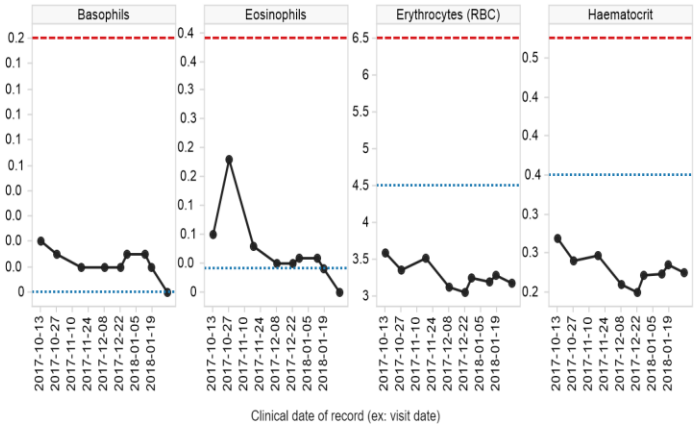
Data Listings for Medical Review

Home | Toggle Zoom Slider | Sort Columns | Reset | Pre-Defined Filters (Click Here) - Liver | Kidney | Heme

ae

Subject	Start Date Int…	Adverse Eve…	Adverse Event	Severity (CT…	Clinical date …
124436	2017-12-18	Nausea	nausea	Grade 1	
124436	2017-11-24	Diarrhoea	Diarrhoea	Grade 2	
124436	2017-11-19	Nausea	nausea	Grade 1	
124436	2017-10-13	Pyrexia	fever	Grade 1	
124436	2017-10-12	Headache	headache	Grade 1	
124436	2017-09-27	Hemiplegia	Nervous syste…	Grade 3	
124436	2017-09-25	Vomiting	vomiting	Grade 1	
124436	2017-09-20	Headache	headaches	Grade 2	
124436	2017-09-19	Nausea	nausea	Grade 1	
124436	2017-09-13	Fatigue	Fatigue	Grade 1	
124438	2018-02-03	Bone pain	wide spead bo…	Grade 3	
124438	2018-02-02	Pulmonary se…	respiratory sep…	Grade 3	
124438	2018-01-27	Musculoskelet…	chest wall pain	Grade 2	
124438	2018-01-23	Musculoskelet…	Chest wall pain	Grade 3	
124438	2017-12-08	Abdominal pain	abdominal pain	Grade 1	
124438	2017-12-08	Fatigue	fatigue	Grade 2	
124438	2017-11-30	Musculoskelet…	Muscoloskelet…	Grade 3	
124438	2017-11-29	Musculoskelet…	Musculoskelet…	Grade 2	
124438	2017-11-29	Nausea	nausea	Grade 2	
124438	2017-11-08	Dyspnoea	Intermittent Dy…	Grade 2	
124438	2017-11-08	Fall	Fall	Grade 2	
124438	2017-11-08	Limb injury	left shoulder in…	Grade 2	
124438	2017-11-08	Pulmonary em…	Pulmonary em…	Grade 3	
124439	2018-01-04	Constipation	Constipation	Grade 2	
124439	2017-12-07	Alanine aminot…	ALT, increased	Grade 4	
124439	2017-12-07	Aspartate ami…	AST, increased	Grade 3	
124439	2017-12-05	Gastric haemo…	Gastric Haem…	Grade 2	
124439	2017-12-01	Constipation	constipation	Grade 1	
124439	2017-11-29	Back pain	lower back pain	Grade 1	
124439	2017-11-06	Viral upper res…	Head Cold	Grade 1	
124440	2018-01-10	Pneumonitis	pneumonitis	Grade 2	
124440	2018-01-06	Abdominal pain	abdominal pain	Grade 2	
124440	2018-01-02	Dry eye	dry eye	Grade 1	
124440	2017-11-29	Fatigue	fatigue	Grade 1	
124442	2018-07-05	Diarrhoea	diarrhoea	Grade 1	
124442	2018-04-26	Nausea	Worsening Na…	Grade 2	
124442	2018-04-25	Vomiting	Vomiting	Grade 1	
124442	2018-04-20	Nausea	Nausea	Grade 1	
124442	2018-04-01	Depression	Depression	Grade 2	

lab



lab

Clinical dat…	AnalyteName	Anal…	La…	LabHigh	LabUnits	LabFlag	LabComments
2018-02-01	Basophils	0.00	0.00	0.20	x10E9/L	0	
2018-02-01	Eosinophils	0.00	0.04	0.44	x10E9/L	-	
2018-02-01	Erythrocytes (RBC)	3.19	4.50	6.50	x10E12/L	-	
2018-02-01	Haematocrit	0.300	0.40	0.54	L/L	-	
2018-02-01	Hemoglobin	104	13…	180.00	g/L	-	
2018-02-01	Leukocytes (WBC)	4.25	3.50	11.00	x10E9/L	0	
2018-02-01	Lymphocytes	0.59	1.50	4.00	x10E9/L	-	

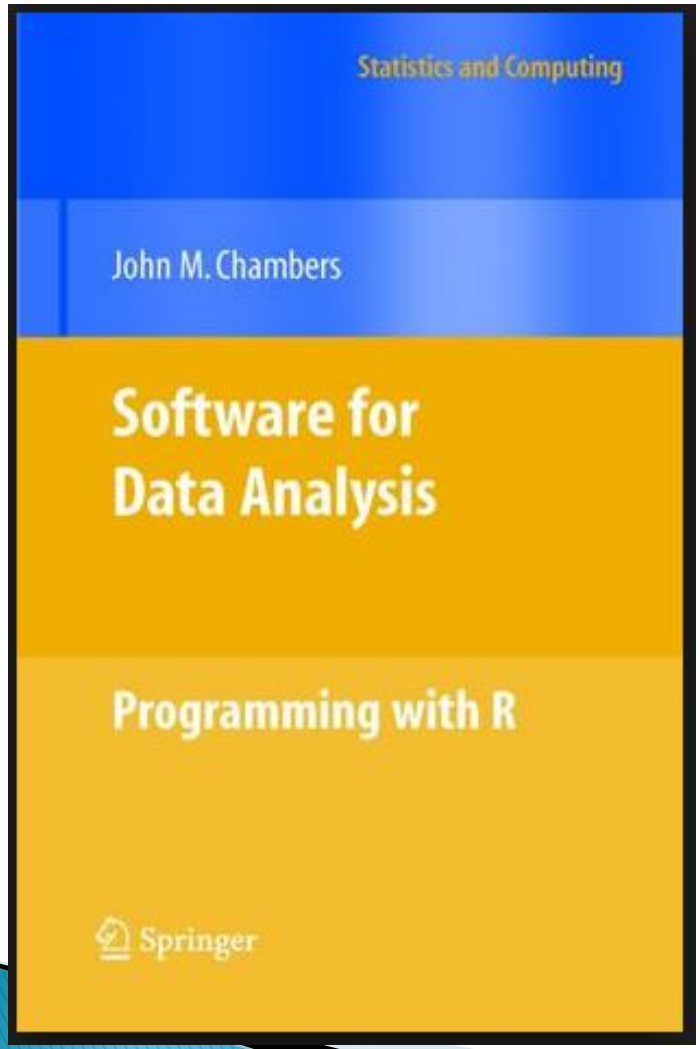
cm

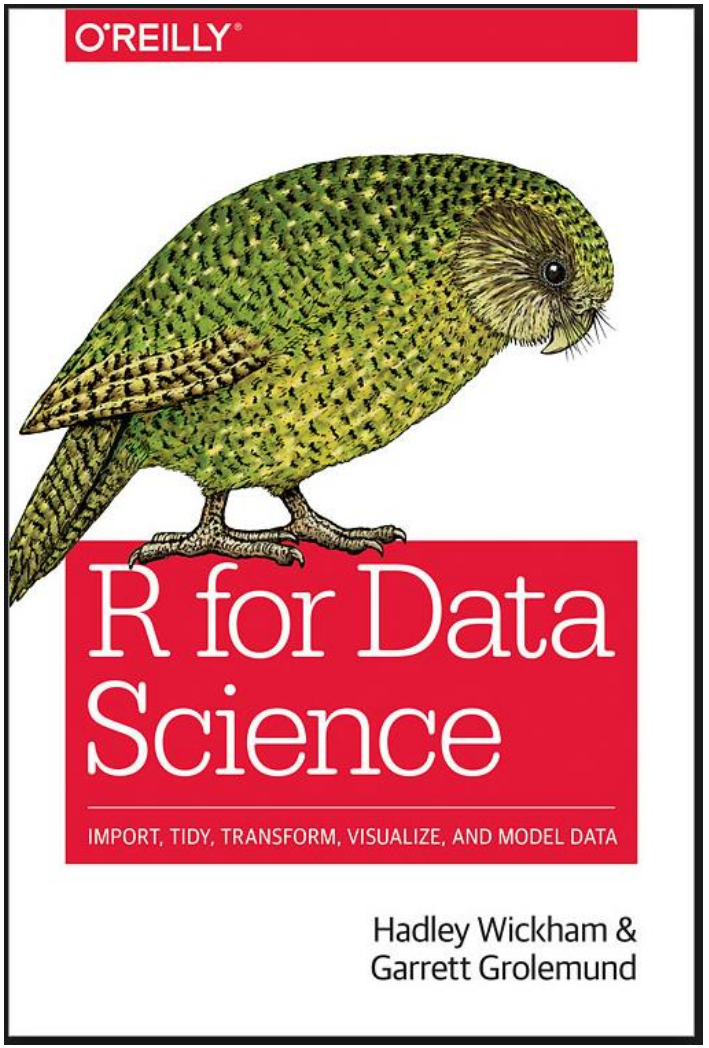
A…	Start Date Int…	Stop Date Int…	Indication	Any Concomi…	Drug Name, …	Drug Name, …	Dose
Yes	2018-02-02		Adverse event	Y	Targin		5/2.5
Yes	2018-01-23		Other	Y	Pantoprazole		40
Yes	2018-01-23		Medical history	Y	Docusate/sen…		50/8
Yes	2017-12-05		Medical history	Y	Florinef (fludro…	FLORINEF	100
Yes	2017-12-01		Adverse event	Y	aspirin (Acetyl…	ASPIRIN /000…	100
Yes	2017-11-24		Other	Y	Zantac	ZANTAC	150
Yes	2017-11-21		Adverse event	Y	clexane	CLEXANE	100
Yes	2017-11-17	12/22/2017 12…	Underlying dis…	Y	Targin		10/5

The R Generation

1996 – Present

Introduction to R





Some R Strengths

- Good user programming interface (Rstudio)
- Extensible: Users can create their own ‘functions’ and libraries
- Create documents with statistical results embeded (R markdown)
- Huge user community
- CRAN: Comprehensive R Archive Network of user function libraries
- Great graphics capabilities and many statistical and scientific packages
- Runs on laptops and scalable to cloud environments
- R Shiny provides a web interface to R functions
- Open source software: Free!

R Weaknesses

- Something new to learn
- Some believe the FDA won't accept R reports
- User is responsible for validation of software
- User CRAN packages generally are not validated

US FDA 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 US 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

phTLGs: R Package for Oncology

Function	Description
ae_sog_tx	Generate Table between system organ class, preferred term and treatment groups
ae_tab_sev_tx	Generate Table between system organ class, preferred terms, severity and treatment groups
ae_tab_tx	Generate a table summary of treatment-emergent adverse events of preferred terms by treatment group
boxplot_change	Boxplot as a function of change over time
boxplot_max	The Box plot against tests maximum values by treatment
boxplot_stats	Boxplots as a function of time along with number graph
boxplot_visits	Boxplots as a function of time

Graphs

Function	Description
ae_dotplot_rr	AE dotplot of incidence and relative risk
boxplot_change	Boxplot as a function of change over time
boxplot_max	The Box plot against tests maximum values by treatment
boxplot_stats	Boxplots as a function of time along with number graph
boxplot_visits	Boxplots as a function of time
forest_plot	Forest plot
swimmer_plot	Swimmer lane plot
tte_plot	A time to event plot
waterfall_plot	In oncology a **special type of bar chart** that illustrates, such details as the response of a number of patients to treatment, the growth of tumors in cancer patients, or the effectiveness of drugs given various criteria. The bars are ordered by outcome of subjects.

R Shiny Swimmer Plot

Tables | Graphs

Customize Graph

Data
swimmerplot_data.csv [Browse...]

Subjects
Subject

Months
Months

Stage
Stage

Continued
Continued

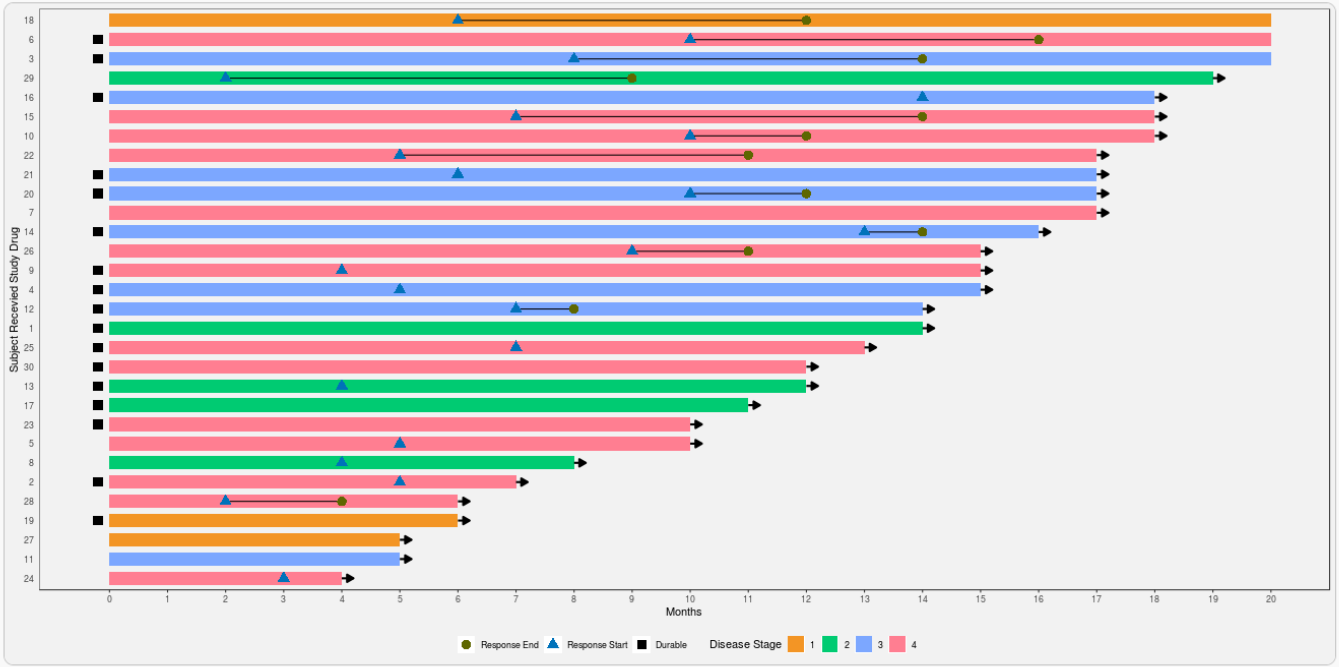
Response Start
RespS

Response End
RespE

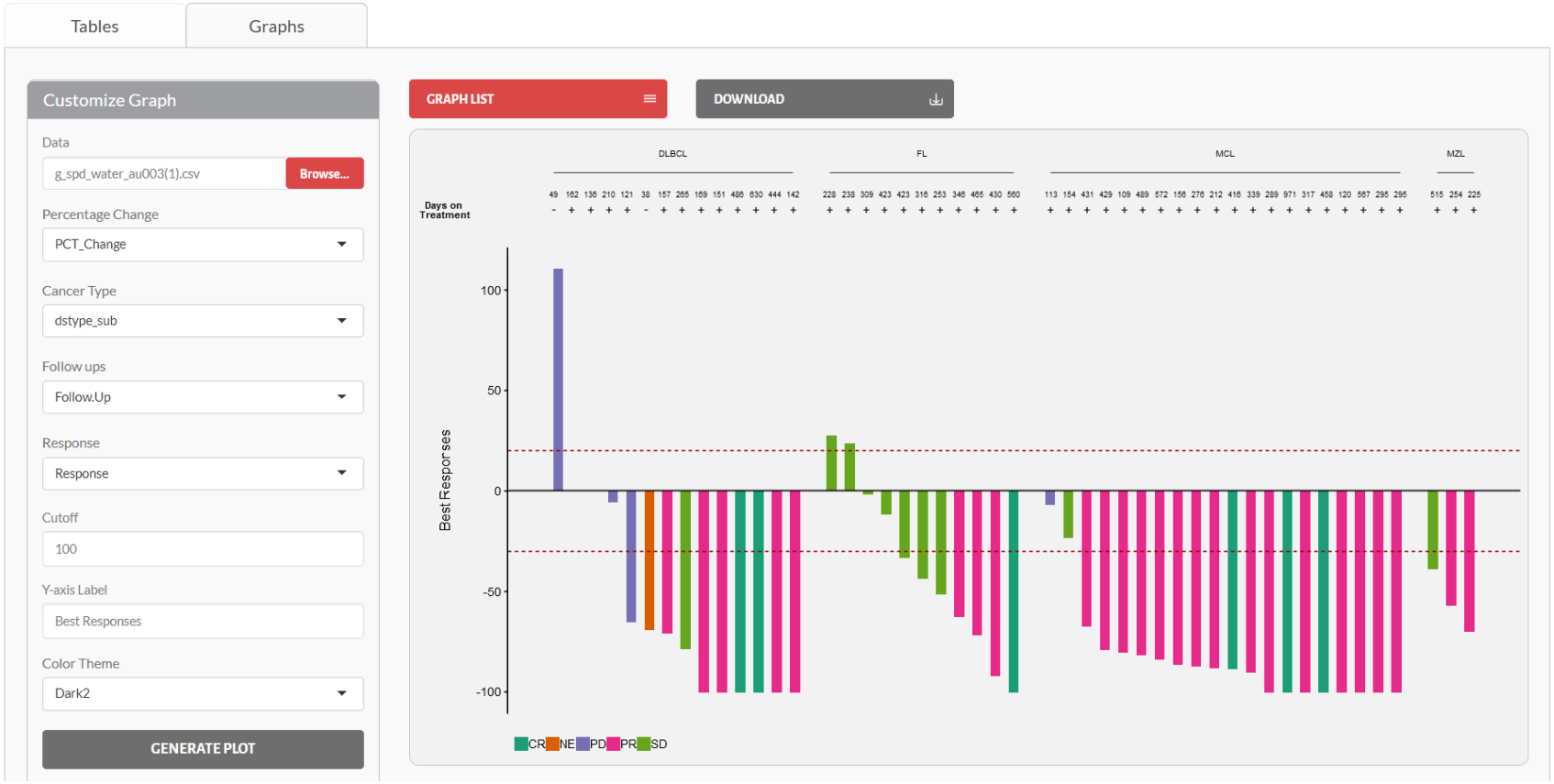
Durable
Durable

GENERATE PLOT

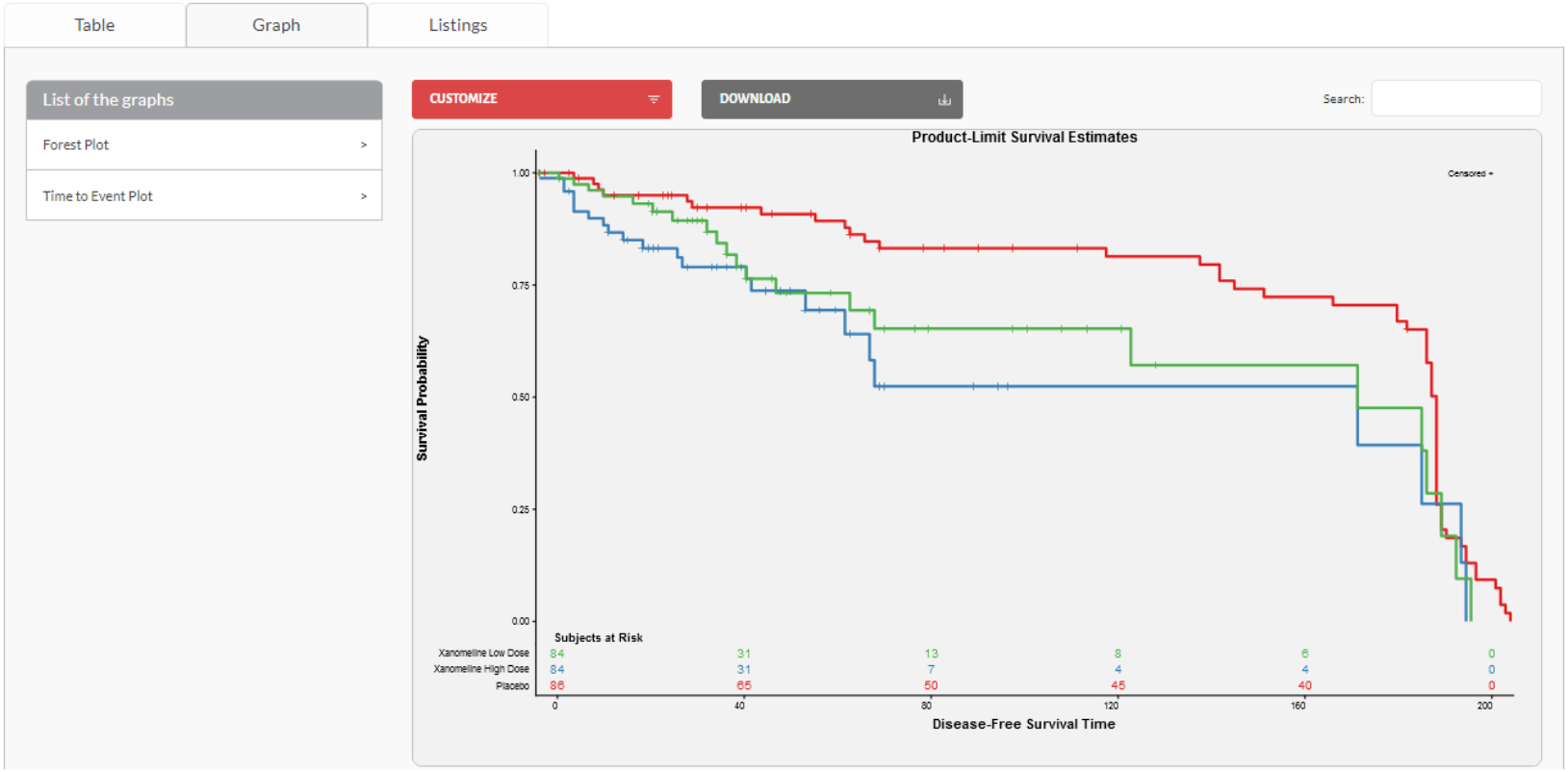
GRAPH LIST | DOWNLOAD



R Shiny Waterfall Plot

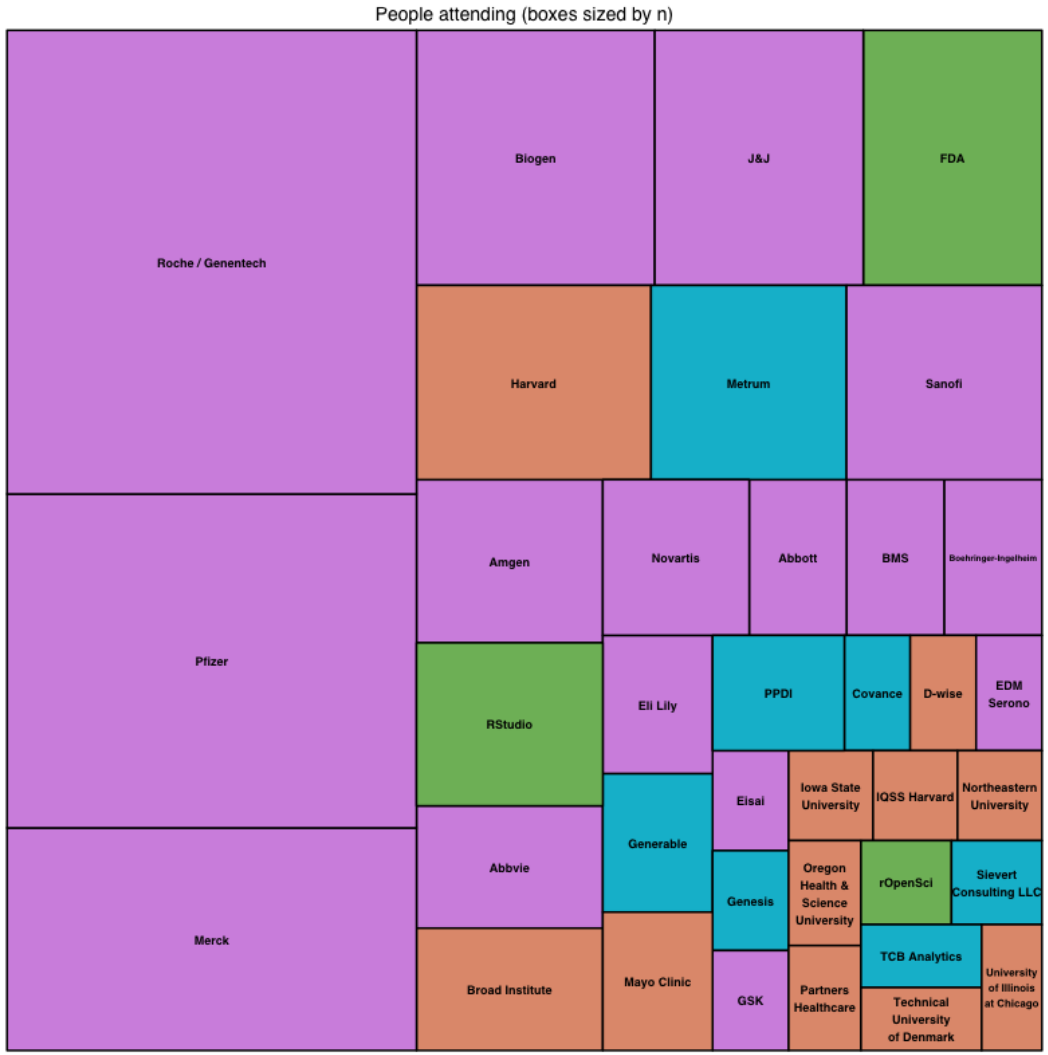


R Shiny Survival Plot



R-consortium-pharma: rinpharma.com

- Distribution of Attendees from R/Pharma 2018
- 15/16 Aug 2018, Harvard University



Final Thoughts

- In less than 2 years we have setup a modern infrastructure to support our many clinical trials.
- The SCE, MDR, and CDR provide a backbone for our data and processes.
- We continue to strive to define business processes in the context of clinical trials operations and regulatory submissions.
- We will evaluate using open-source software to take advantage of numerous opportunities in our business.

Contact Information

Alan Hopkins, Ph.D.
BeiGene Ltd.
1900 Powell St., Suite 500
Emeryville, CA 94608
Email: alan.hopkins@beigene.com
Web: www.beigene.com

Acknowledgments: Colleagues at BeiGene have been helpful with the software concepts: Debi Roy, Michael Hagendoorn, Chrissie Neumann, Sammy Zhang and Josh Patel.