Scientific Computing and Consulting
Novartis Pharmaceuticals

Would John, Paul, George or Ringo have been famous if it were not for The Beatles?

Douglas M. Robinson, PhD
Global Head, Scientific Computing and Consulting
Advanced Methodology and Data Science
Novartis Pharmaceuticals

NOVARTIS | Reimagining Medicine

Disclaimer

- This presentation is based on publicly available information (including data relating to non-Novartis products or approaches)
- Any data relating to Novartis products have been either anonymized or simulated to represent actual data
- The views presented are the views of the presenter, not necessarily those of Novartis
- These slides are intended for educational purposes only and for the personal use of the audience
- The content of this slide deck is accurate to the best of the presenter's knowledge at the time of production

Background on title…

- Personal silver lining to the pandemic – spending more time with family
  - Taking walks around neighborhood talking and listening to music
- Daughter developed a healthy obsession with The Beatles

Motivation: Why things have to change

- Teams pre-plan 100's of Tables, Listings in studies
  - Regardless of value
  - Static outputs in SAS
- Generate 1000s of formatted output for submission dossiers plus 1000s more for Advisory Committees

Challenge Statement:

- We try to cover all possibilities
  - Requires heavy resources, yet many (most) not even referenced in reports
- If trials show unexpected results...
  - Must produce **even more** output

What are we supposed to do?

- Earn the trust of our collaborators
  - Understand scientific question: disease, mechanism of action, technology
  - Consider the data: identify sources of uncertainty
  - Listen carefully, ask questions and **really think**
- "There is no single statistical tool that is as powerful as a well-chosen graph[1]"
  - Effective visualizations **communicate complex information** with ease
  - Facilitate insight, understanding and **decision making**
- We formed the Advanced Visual Analytics Initiative:
  - Promote **graphical thinking**
  - Use **good graphical principles**
  - Leverage **R – Shiny Applications** to **dynamically explore** our data

1. Chambers, J.M., Cleveland, W.S., Kleiner, B. & Tukey, P.A. Graphical Methods for Data Analysis (Wadsworth, Pacific Grove, CA, 1983)

Introducing the Graphical Principles Cheat Sheet!!

Download free at https://graphicsprinciples.github.io/

Graphics Tutorial: https://ascpt.onlinelibrary.wiley.com/doi/full/10.1002/psp4.12455

Authors: Alison Margolskee, Mark Baillie, Baldur Magnusson, Julie Jones, Marc Vandemeulebroecke

Communication

Effective visualizations communicate complex statistical and quantitative information facilitating insight, understanding, and decision making.

But what is an effective graph?

This cheat sheet provides general guidance and points to consider.

Planning

- **Why** – Clearly identify the purpose of the graph, e.g. to deliver a message or for exploration?
- **What** – Identify the quantitative evidence to support the purpose
- **Who** – Identify the intended audience (specialists, non-specialists, both) and focus the design to support their needs
- **Where** – Adapt the design to space or formatting constraints (e.g. clinical report, slide deck or publication)

Principles of Effective Graphic Design

Proximity – group related elements together

Alignment – elements on the same vertical or horizontal plane are perceived as having similar properties

Simplicity – cut anything superfluous, only include elements that add value, limit to 2-3 colors or fonts

White space (empty space) – use white space to minimize distraction & provide clarity

Legibility – sans serif fonts are easier to read, use color for emphasis instead of a new typeface

Color – select colors that present enough contrast to make the graph legible. Choose monochromatic color schemes to prevent clashing. Use dark colors and accent colors to emphasize important information

Visual Hierarchy – use color, font, image size, typeface, alignment & placement to create a viewing order

Focal Points – primary area of interest that immediately attracts the eye, emphasize the most important concept and make it your focal point. Use contrasting colors to draw attention

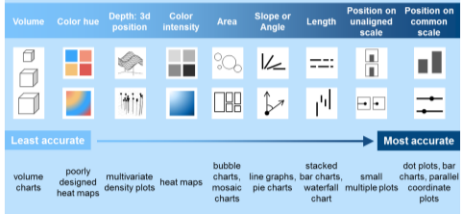
Repetition – repeating elements can be visually appealing, repeated shapes, labels, colors

Familiarity – using familiar styles, icons, navigation structure makes viewers feel confident

Consistency – be consistent with heading sizes, font choices, color scheme, and spacing. Use images with similar styles

Effectiveness Ranking

A graph is a representation of data that visually encodes numerical values into attributes such as lines, symbols and colors. The Cleveland-McGill scale can be used to select the most effective attribute(s) for your purpose.



Selecting the right base graph

Consider if a standard graph can be used by identifying suitable designs based on the:
(i) **purpose** (i.e. message to be conveyed or question to answer) and (ii) **data** (i.e. variables to display).

Example plots categorized by purpose

Deviation	Correlation	Ranking	Distribution	Evolution	Part-to-whole	Magnitude
Chg. from baseline	Scatter plot	Horizontal bar chart	Boxplot	Kaplan Meier	Stacked bar chart	Vertical bar chart
Waterfall	Heat map	Dotplot	Histogram	Line plot	Tree map	Forest plot

Utilize the **Graph Gallery** (go/graphgallery). There are many effective base designs that can be adapted to your purpose.

Facilitating Comparisons

Proximity improves association

Place labels next to data instead of using legends

Group together elements to be compared directly

Ease visual inspection

Order values to help compare across many categories

Judgments are easier to make on a common vertical scale

Reduce mental arithmetic

Plot the final comparison e.g. mean difference not two means
Exception: if comparator is of interest in itself

Use reference lines and other visual anchors.

Color for emphasis or distinction

Restrained use of color is highly effective in organizing a narrative and calling attention to certain elements.

Think carefully before introducing additional color. Do you really need it?

Do not use color to differentiate between categories of the same variable

Use colors or shades to represent meaningful differences such as positive/negative values, treatments or doses

Be consistent, use the same color to mean the same thing in a series of graphs (e.g. treatment, dose)

Use a bold, saturated or contrasting color to emphasize important details.

Emphasize the data by minimizing unnecessary ink, e.g. soften gridlines with a light color

Utilize existing resources for selection of appropriate palettes such as Color brewer or Munsell

Implementation Considerations

Plot cause on the x-axis and effect on the y-axis. Use this standard convention in order to avoid misinterpretation.

Aspect ratio can influence interpretation. Aim for a 45 degree angle of change to avoid over-interpretation of slope.

Use position for comparisons rather than length (i.e. dots instead of bars), especially for non-linear scales (e.g. log scale or % change).

Do not plot log-normally distributed variables on a linear scale (e.g. hazard ratio, AUC, CL)

When displaying data measured on the same scale, also plot them on the same scale for easy comparison.

Connected data imply continuity. Do not connect data across a disconnected or uneven time scale.

Visits displayed close together are perceived to be closer in time. Space the visits proportional to the time between each in order to avoid confusion.
Exception: baseline or pre-dose

Plot data and inferences to support stories about models.

Legibility and Clarity

Effective graphs stand alone. They use titles, annotations, labels, shapes, colors, and textures to deliver important information.

Label axes with clear measurement units and provide annotations that support the message.

Use font size to create hierarchy (e.g. set titles 2pt larger than all other labels to make them more prominent).

Do not type too small or too condensed. Break long titles into two lines. Shift or adjust size of labels that overlap.

Keep the font style simple, sans serif is easier to read.

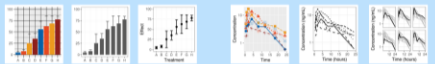
Display text with enough contrast to be visible. Favor the use of dark on light instead of light on dark whenever possible.

Bold or italics should only be used for layering or emphasis. Emphasizing everything means nothing gets emphasized.

Try not to set text at an angle, as this decreases readability. Think of alternative solutions such as transposing the graph.

Putting it all together – Remove the clutter & emphasize the message

Creating a graph is an iterative process: produce, review and refine.



Colors, backgrounds, and borders can be removed and gridlines reduced.

It is easier to see differences in position over a difference in length, i.e. a dot over a bar.

Using too many colors can be distracting. Use white background and try using other methods to distinguish different curves.

One solution could be repeating the data in different panels, highlighting individual curves in a darker color.

Good graph checklist

Clear Communication

- ❑ Is the message of the graph as clear as possible?
- ❑ Is it easy for someone unfamiliar with the data to interpret the graph?
- ❑ Are the patterns/relationships easily identified?
- ❑ Is the graph tailored to its primary purpose and audience?
- ❑ Is the correct graph type used?

Facilitating Comparisons

- ❑ Are elements to be compared grouped together?
- ❑ Are labels placed next to data instead of in legends?
- ❑ Have categories been ordered for easy comparison?
- ❑ Can the plot be read without doing mental calculations?
- ❑ Are the estimates of interest plotted (e.g. mean differences with confidence intervals)?

Color for emphasis or distinction

- ❑ Are graphical elements displayed in a dark color on a light background?
- ❑ Are grid lines drawn with a thin line and a light color such as grey?
- ❑ Are colors used sparingly (e.g. max 3)?
- ❑ Do all elements in the graph have a purpose (e.g. colors, textures, grid lines)?
- ❑ Are the same colors used to mean the same thing in a series of graphs?

Implementation Considerations

- ❑ Are multiple panels plotted on the same scale?
- ❑ Are lognormally distributed variables plotted on a log scale?
- ❑ Are common baselines used wherever possible?
- ❑ Does the orientation of the axes aid interpretation?
- ❑ Does the aspect ratio allow the reader to see variations in the data?
- ❑ Are data across a disconnected time scale kept disconnected?
- ❑ Are data spaced proportionally to the actual time interval (instead of according to visit number)?
- ❑ Are data and inferences plotted to support stories about models?
- ❑ Are number of patients by group reported if this adds context?

Legibility and Clarity

- ❑ Can all graphical elements be seen?
- ❑ Does the graph have a clear title, axis labels, annotations and data units?
- ❑ Can the font be read without eye strain or effort?
- ❑ Are sans-serif fonts used?
- ❑ Do text sizes have correct hierarchy (big to small, main text to subtext)?
- ❑ Are the elements of the graph clearly labeled (e.g. points, error bars, lines, shaded regions)?
- ❑ Are labels oriented horizontally where possible?

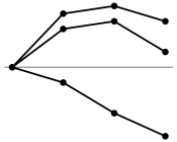
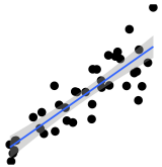
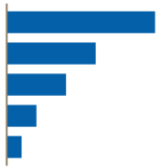
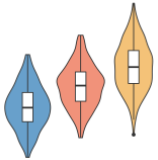
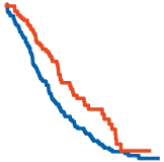
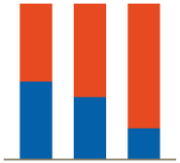
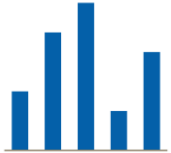
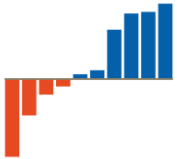
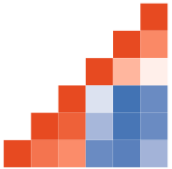
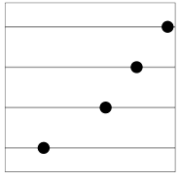
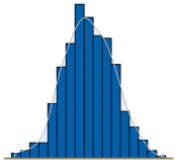
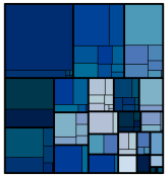
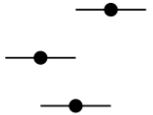
Resources

Books:
E. R. Tufte, The visual display of quantitative information, Connecticut, Graphics. Press, 2001.
Cleveland, W.S. and McGill, Robert, Graphical perception: theory, experimentation and application to the development of graphical methods. JASA, Vol. 79, No. 387, pp. 531—554, 1984.
S. Few, Show Me The Numbers - Designing Tables and Graphs to Enlighten (2nd. Edition), Burlingame, CA: Analytics Press, 2012.
D. M. Wong, The Wall Street Journal Guide to Information Graphics: The Dos and Don'ts of Presenting Data, Facts, and Figures. December 16, 2013.
J. Doumont, Trees, maps, and theorems: Effective communication for rational minds. PRINCIPIAE.
N. B. Robbins, Creating More Effective Graphs. Chart House.

Online resources:
https://www.perceptualedge.com/ (S. Few)
https://www.edwardtufte.com/tufte/ (E. Tufte)
http://flowingdata.com/ (N. Yau)
http://www.principiae.be/ (J. Doumont)
http://andrewgelman.com/ (A. Gelman)
http://www.thefunctionalart.com/ (A. Cairo)
http://www.nbr-graphs.com/ (N. Robbins)

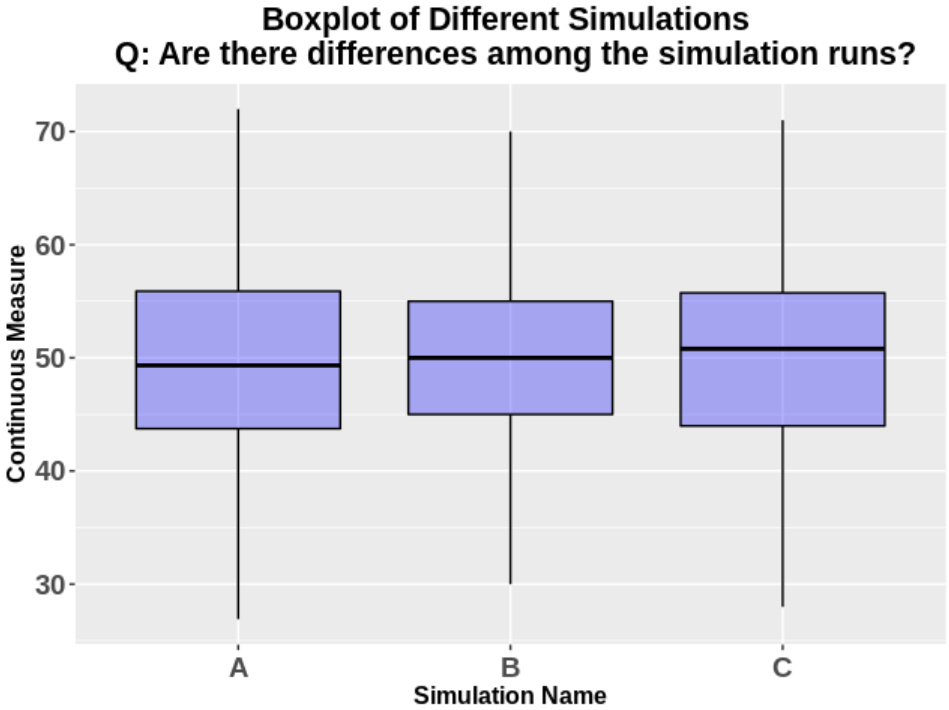
NOVARTIS | Reimagining Medicine

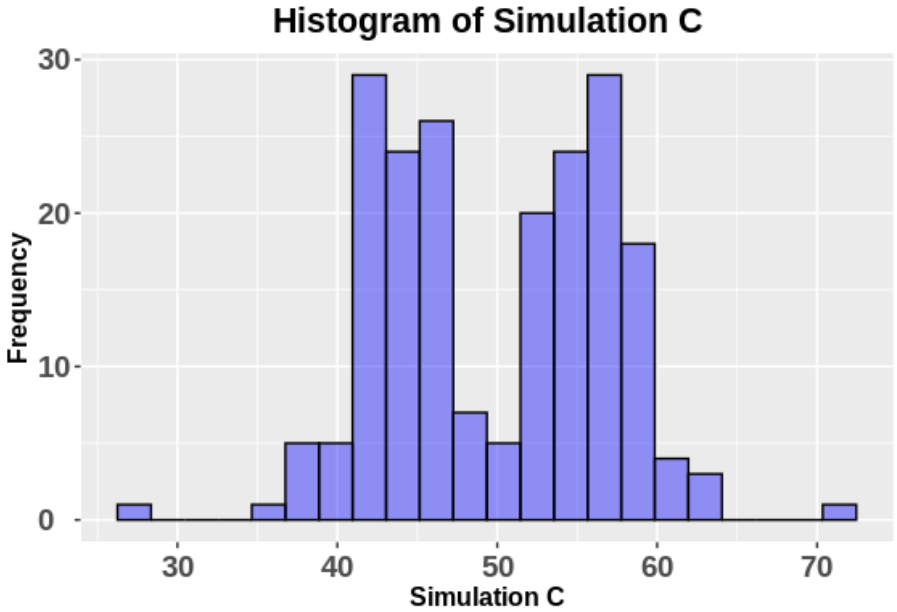
Choosing the Correct Graph Type Aids Interpretation

Deviation	Correlation	Ranking	Distribution	Evolution	Part-to-whole	Magnitude
Chg. from baseline	Scatter plot	Horizontal bar chart	Boxplot	Kaplan Meier	Stacked bar chart	Vertical bar chart
						
Waterfall	Heat map	Dotplot	Histogram	Line plot	Tree map	Forest plot
						

- Carefully consider your question and its purpose; choose wisely
- Combine graphs with statistical thinking to may informed decisions
- Strive for simplicity to maximize the value of your data

Is it enough to just use graphics?





- It is <u>not enough</u> to “just use graphics”
- Experiment to allow your data to speak

Potential impact on Data Monitoring Committees (DMC)

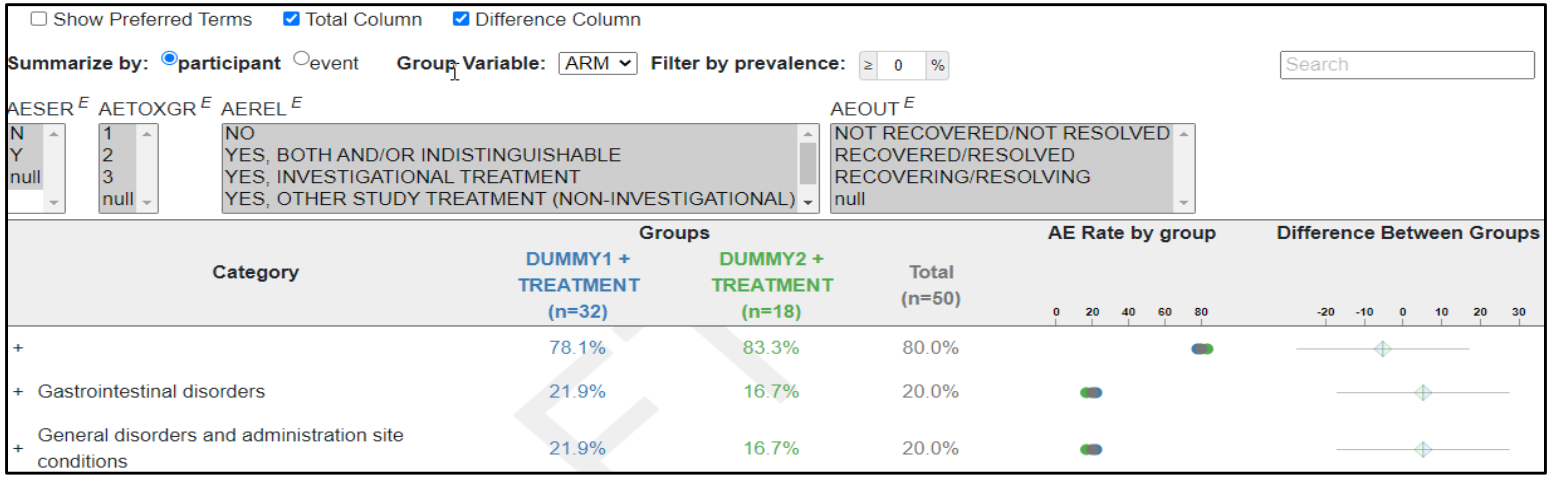
- Tim Friede is a Professor of Biostatistics at the University Medical Center Göttingen, Germany
- Independent statistician on numerous Data Monitoring Committees
  - 15 years of DMC experience, > 50 DMC's!
  - Pre-print of journal article DMC for Covid-19
- Comments:
  - DMCs sometimes get 1000's of pages of output to review
    - Mostly tables and listings, rarely figures
  - Difficult to see data from different data domains
  - Some DMCs had 2019 data cut for June 2020 meeting
  - Data questions typically addressed in subsequent DMC mtg
    - Want to see the answer > if nothing, then move on
  - Want ability to explore *ad hoc* subgroups

First qualified Shiny App used for Novartis DMC!

- In Feb 2020, Novartis started working with contractors from Rho, Inc.
  - Stemmed from close collaboration on Interactive Safety Graphics (ISG)
    - Subteam of the ASA Biopharm-DIA Safety Working Group
    - https://safetygraphics.github.io
  - Expansion beyond eDISH to safety explorer suite
- Incorporated safety explorer into Novartis exploration workflow
- Novartis initiated multiple trials on COVID-19 with single DMC
  - Shiny app to aid DMC safety data review
  - Weekly updates
- Tim Friede truly enjoyed experience!
  - DMC members highly engaged and able to quickly identify required information
  - Requested additional functionality to aid review
- How to use apps for fully external DMCs?

Generalized tools: Adverse Event Explorer Module

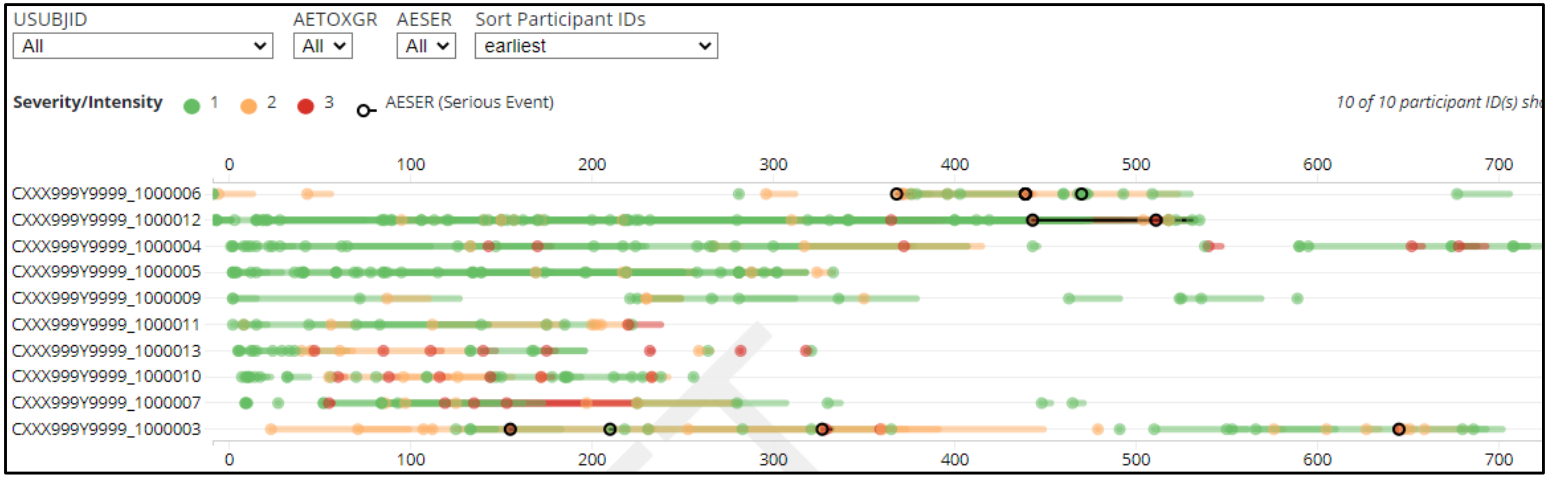
Based on Safety Explorer Suite https://rhoinc.github.io/safety-explorer-suite/docs/



- Provides Adverse Events (AE) categorized by System Organ Class (SOC)
  - AE rates by treatment and overall (Numerically and in Dot plot)
  - Rate comparison forest plot
- Clicking “+” breaks down the SOC by preferred term
- Powerful search box: “head” searches for headache and tension headache

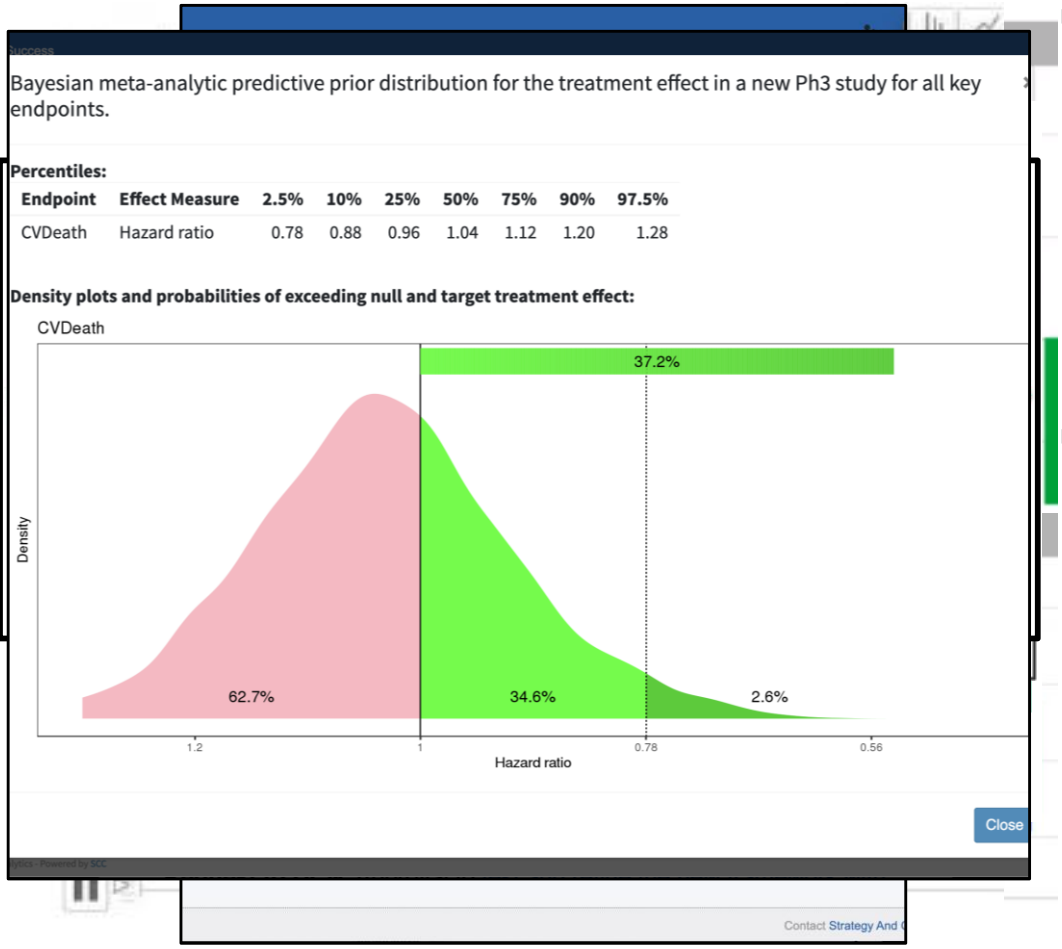
Generalized tools: Adverse Event Timelines Module

Based on Safety Explorer Suite https://rhoinc.github.io/safety-explorer-suite/docs/



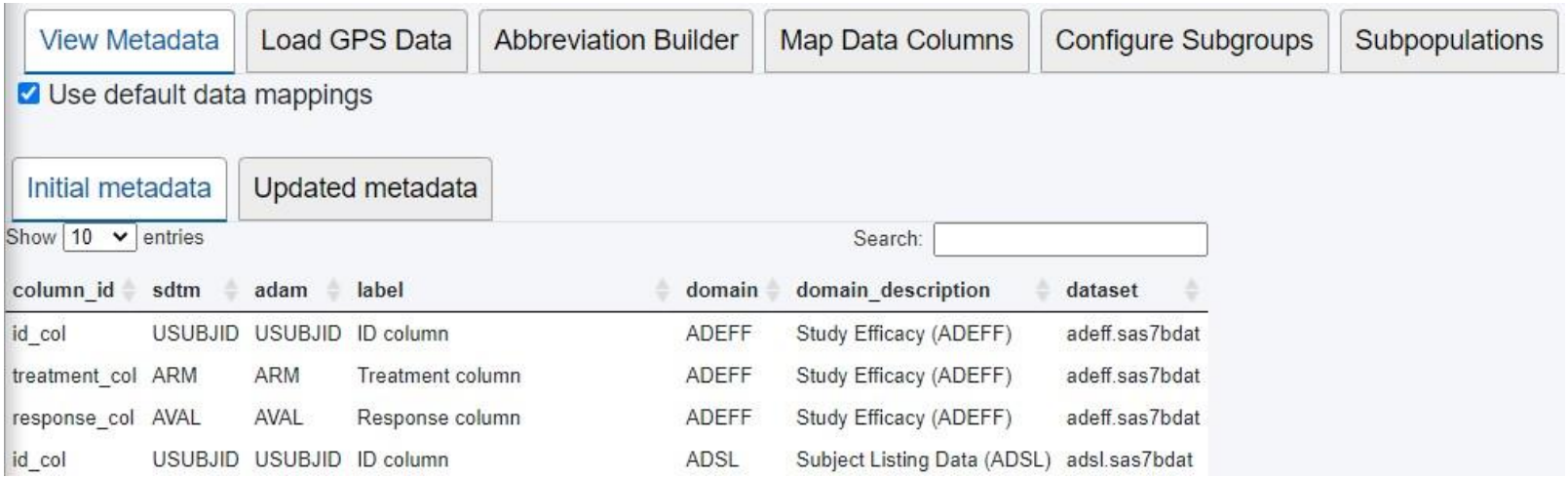
- Adverse Event (AE) swim lane colored by severity
  - Patients sorted by time of earliest adverse event
- Serious adverse events are easily identified by black circles
  - Hovering over provides AE term, start and end dates
- Individual AE chart per patient allows focused exploration

Supporting clinical projects and non-clinical projects



- SCC supports many clinical projects
  - Study specific applications
  - Ideal breeding ground for new ideas
  - See what resonates with our teams
  - See what may be generalized and used across programs
- Non-clinical projects
  - ARV: Analytics Resource Visualization
    - Custom built database and visualization tool
  - Probability of success
    - Interface to a new framework that provides a more robust and reliable PoS estimates

Harmonized Shiny Framework



column_id	sdtm	adam	label	domain	domain_description	dataset
id_col	USUBJID	USUBJID	ID column	ADEFF	Study Efficacy (ADEFF)	adeff.sas7bdat
treatment_col	ARM	ARM	Treatment column	ADEFF	Study Efficacy (ADEFF)	adeff.sas7bdat
response_col	AVAL	AVAL	Response column	ADEFF	Study Efficacy (ADEFF)	adeff.sas7bdat
id_col	USUBJID	USUBJID	ID column	ADSL	Subject Listing Data (ADSL)	adsl.sas7bdat

- Challenge: Heterogeneous app development styles + reinventing repetitive tasks
- **Harmonized framework** to streamline app development
- We do "**the dirty work**": accessing data, mapping data to charts, configure subgroups and subpopulations, generate label abbreviations
- Consistent modules that are **plug and play** are key to scalability goals

We are not done…

- How to handle large data size
  - Lab data from cardiovascular “mega trials”, AE data from large Onc studies
  - Exploring data from pooled studies
  - New data modalities: imaging, digital sensors
- Animations
  - Packages to visualize temporal data, patient journeys, etc.
- Streamline the qualification process
  - Adapt current process where traditional qualification methods break down
- Increased awareness + education
  - Many teams have still not experienced benefits : Seeing is believing

Conclusion

- Focus on good graphical principles
  - **Graphic Principles Cheat Sheet**: https://graphicsprinciples.github.io/
- Harness the potential of data through graphics and statistical thinking
- Shiny is a **powerful tool** to engage clinical teams
  - Promote learning from data
  - Reduce down the static tables and listings
- Leverage each other's strengths and innovations
  - See what works and generalize
- Engaged teams + Senior Management support + Analysts having fun =
  - **The ultimate win – win!!**

But wait…

Would John, Paul, George or Ringo have been famous if it were not for The Beatles?

…we may never know

but I do know that for me in drug development

“I get by with a little help from my friends”

Friends that have helped me get by

Scientific Computing and Consulting
David Granjon
David Hall
Mustapha Larbaoui
Ardalan Mirshani
Bo Wang
Flavio Lombardo (intern)

Advanced Visual Analytics
Bibiana Blatna
Allison Florance
Joseph Kahn
Alison Margolskee
Craig Wang
Andrew Wright

Effective Visual Communication
Mark Baillie
Julie Jones
Baldur Magnusson
Marc Vandemeulebroecke

Rho, Inc.
Natalia Andriychuk
Preston Burns
Ryan Chronowic
Becca Krouse

Gilead Sciences
Jeremy Wildfire

Serapta Therapeutics
Xiao Ni

Novartis Analytics
Mouna Akatcha
Jahangir Alam
Janice Branson
Spencer Childress
Tuochuan Dong
Eric Gibson
Kaustav Nandy
David Ohlssen
Xu Shu

University Medical Center Göttingen
Tim Friede

Daughter
Samantha Robinson

NOVARTIS | Reimagining Medicine

Goo Goo G'Joob!

Douglas M. Robinson, PhD

Global Head, Scientific Computing and Consulting

Advanced Methodology and Data Science

Novartis Pharmaceuticals

douglas.robinson@novartis.com