


# How do we make better graphs? Effective visual communication for the quantitative scientist

**Mark Baillie**  
June 6<sup>th</sup>, 2019

**ARTICLE**    **OPEN**

# Scalable and accurate deep learning with electronic health records

Alvin Rajkomar <sup>1,2</sup>, Eyal Oren<sup>1</sup>, Kai Chen<sup>1</sup>, Andrew M. Dai<sup>1</sup>, Nissan Hajaj<sup>1</sup>, Michaela Hardt<sup>1</sup>, Peter J. Liu<sup>1</sup>, Xiaobing Liu<sup>1</sup>, Jake Marcus<sup>1</sup>, Mimi Sun<sup>1</sup>, Patrik Sundberg<sup>1</sup>, Hector Yee<sup>1</sup>, Kun Zhang<sup>1</sup>, Yi Zhang<sup>1</sup>, Gerardo Flores<sup>1</sup>, Gavin E. Duggan<sup>1</sup>, Jamie Irvine<sup>1</sup>, Quoc Le<sup>1</sup>, Kurt Litsch<sup>1</sup>, Alexander Mossin<sup>1</sup>, Justin Tansuwan<sup>1</sup>, De Wang<sup>1</sup>, James Wexler<sup>1</sup>, Jimbo Wilson<sup>1</sup>, Dana Ludwig<sup>2</sup>, Samuel L. Volchenboum<sup>3</sup>, Katherine Chou<sup>1</sup>, Michael Pearson<sup>1</sup>, Srinivasan Madabushi<sup>1</sup>, Nigam H. Shah<sup>4</sup>, Atul J. Butte<sup>2</sup>, Michael D. Howell<sup>1</sup>, Claire Cui<sup>1</sup>, Greg S. Corrado<sup>1</sup> and Jeffrey Dean<sup>1</sup>

## ARTICLE OPEN

# Scalable and accurate deep learning with electronic health records

Alvin Rajkomar<sup>1,2</sup>, Eyal Oren<sup>1</sup>, Kai Chen<sup>1</sup>, Andrew M. Dai<sup>1</sup>, Nissan Hajaj<sup>1</sup>, Michaela Hardt<sup>1</sup>, Peter J. Liu<sup>1</sup>, Xiaobing Liu<sup>1</sup>, Jake Marcus<sup>1</sup>, Mimi Sun<sup>1</sup>, Patrik Sundberg<sup>1</sup>, Hector Yee<sup>1</sup>, Kun Zhang<sup>1</sup>, Yi Zhang<sup>1</sup>, Gerardo Flores<sup>1</sup>, Gavin E. Duggan<sup>1</sup>, Jamie Irvine<sup>1</sup>, Quoc Le<sup>1</sup>, Kurt Litsch<sup>1</sup>, Alexander Mossin<sup>1</sup>, Justin Tansuwan<sup>1</sup>, De Wang<sup>1</sup>, James Wexler<sup>1</sup>, Jimbo Wilson<sup>1</sup>, Dana Ludwig<sup>2</sup>, Samuel L. Volchenboum<sup>3</sup>, Katherine Chou<sup>1</sup>, Michael Pearson<sup>1</sup>, Srinivasan Madabushi<sup>1</sup>, Nigam H. Shah<sup>4</sup>, Atul J. Butte<sup>2</sup>, Michael D. Howell<sup>1</sup>, Claire Cui<sup>1</sup>, Greg S. Corrado<sup>1</sup> and Jeffrey Dean<sup>1</sup>

Scalable and accurate deep learning with electronic health records ...

<https://www.nature.com> > [npj digital medicine](#) > [articles](#)

by A Rajkomar - 2018 - [Cited by 163](#) - [Related articles](#)

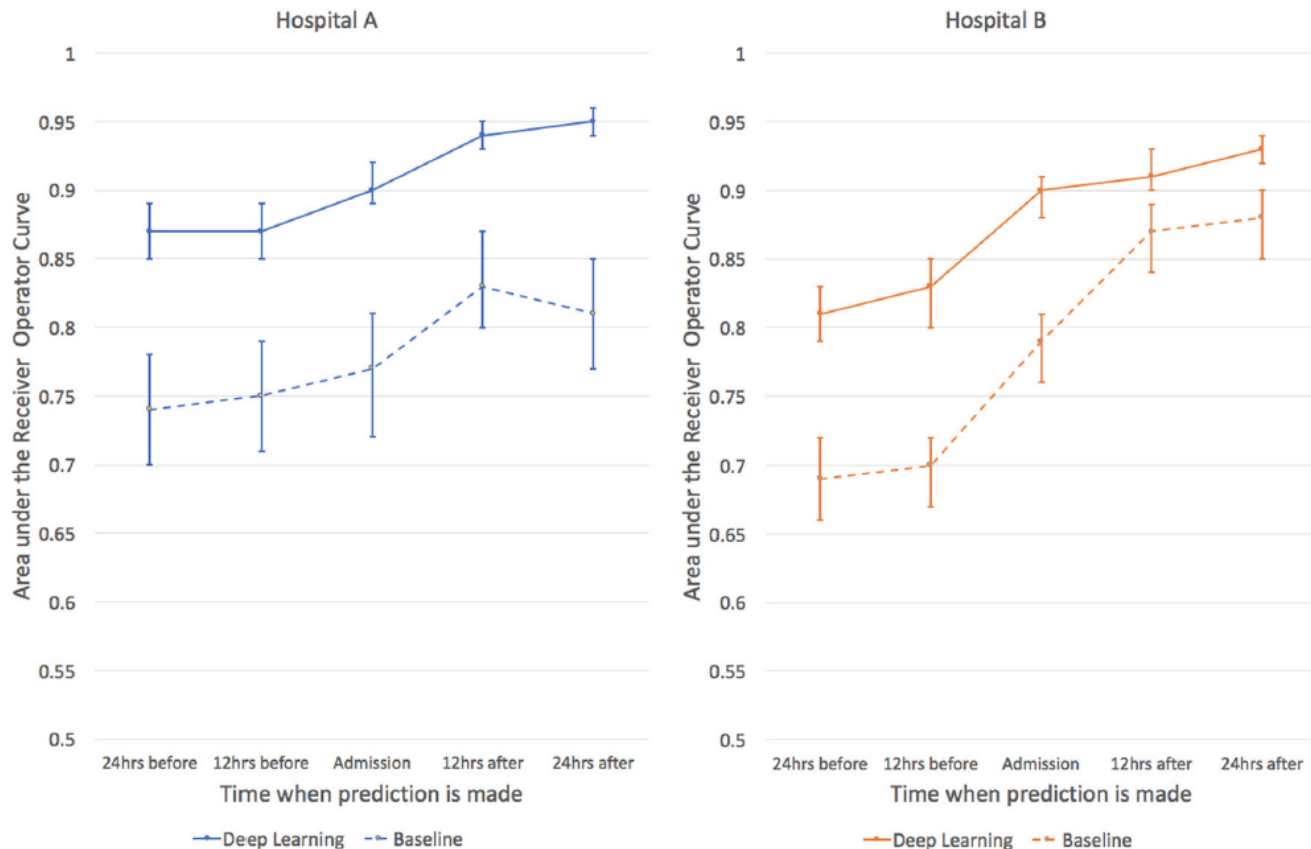
May 8, 2018 - Predictive modeling with electronic health record (EHR) data is anticipated to drive personalized medicine and improve healthcare quality.

You've visited this page 5 times. Last visit: 5/18/19

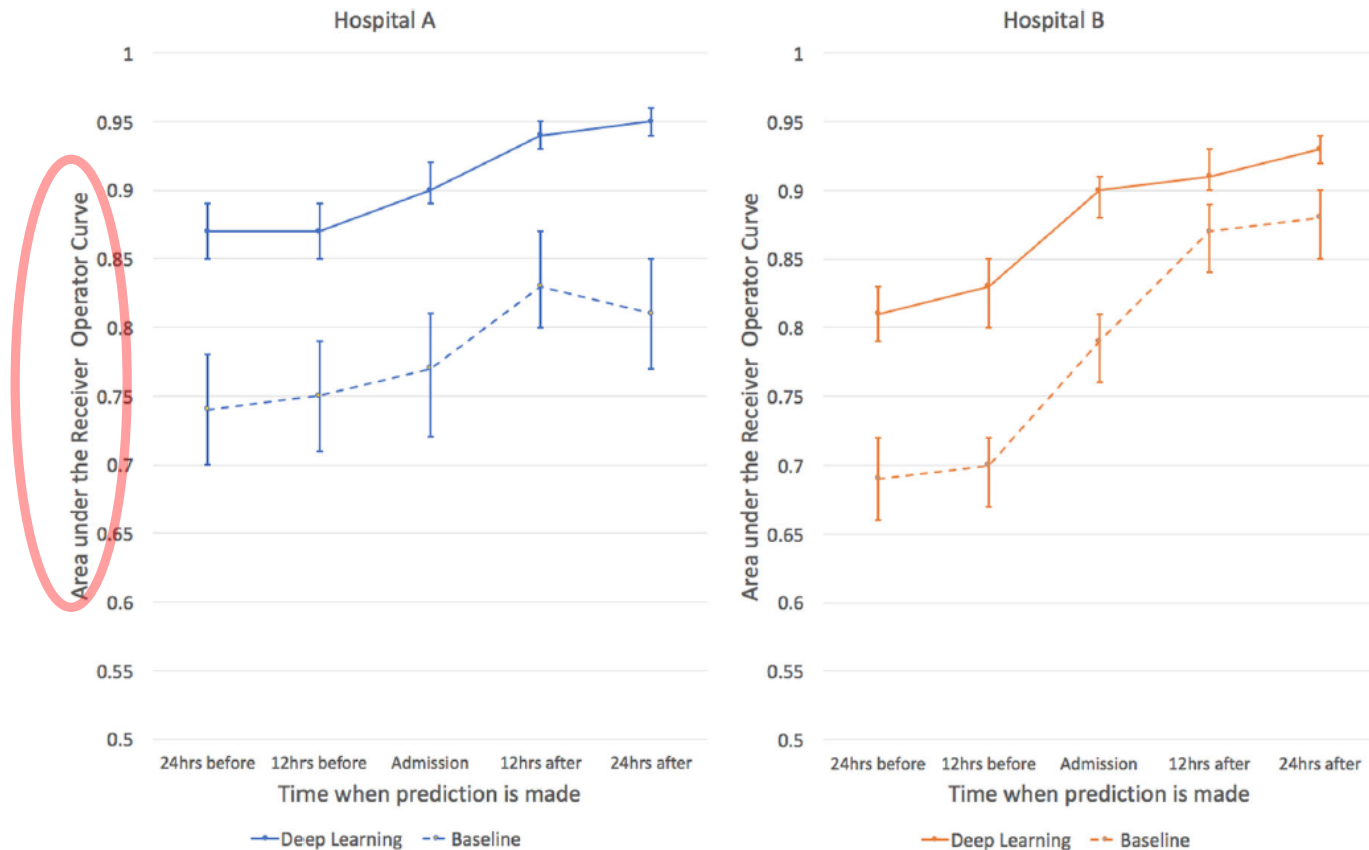
# Scalable and accurate deep learning with electronic health records

Alvin Rajkomar<sup>1,2</sup>, Eyal Oren<sup>1</sup>, Kai Chen<sup>1</sup>, Andrew M. Dai<sup>1</sup>, Nissan Hajaj<sup>1</sup>, Michaela Hardt<sup>1</sup>, Peter J. Liu<sup>1</sup>, Xiaobing Liu<sup>1</sup>, Jake Marcus<sup>1</sup>, Mimi Sun<sup>1</sup>, Patrik Sundberg<sup>1</sup>, Hector Yee<sup>1</sup>, Kun Zhang<sup>1</sup>, Yi Zhang<sup>1</sup>, Gerardo Flores<sup>1</sup>, Gavin E. Duggan<sup>1</sup>, Jamie Irvine<sup>1</sup>, Quoc Le<sup>1</sup>, Kurt Litsch<sup>1</sup>, Alexander Mossin<sup>1</sup>, Justin Tansuwan<sup>1</sup>, De Wang<sup>1</sup>, James Wexler<sup>1</sup>, Jimbo Wilson<sup>1</sup>, Dana Ludwig<sup>2</sup>, Samuel L. Volchenboum<sup>3</sup>, Katherine Chou<sup>1</sup>, Michael Pearson<sup>1</sup>, Srinivasan Madabushi<sup>1</sup>, Nigam H. Shah<sup>4</sup>, Atul J. Butte<sup>2</sup>, Michael D. Howell<sup>1</sup>, Claire Cui<sup>1</sup>, Greg S. Corrado<sup>1</sup> and Jeffrey Dean<sup>1</sup>

Predictive modeling with electronic health record (EHR) data is anticipated to drive personalized medicine and improve healthcare quality. Constructing predictive statistical models typically requires extraction of curated predictor variables from normalized EHR data, a labor-intensive process that discards the vast majority of information in each patient's record. We propose a representation of patients' entire raw EHR records based on the Fast Healthcare Interoperability Resources (FHIR) format. We demonstrate that deep learning methods using this representation are capable of accurately predicting multiple medical events from multiple centers without site-specific data harmonization. We validated our approach using de-identified EHR data from two US academic medical centers with 216,221 adult patients hospitalized for at least 24 h. In the sequential format we propose, this volume of EHR data unrolled into a total of 46,864,534,945 data points, including clinical notes. Deep learning models achieved high accuracy for tasks such as predicting: in-hospital mortality (area under the receiver operator curve [AUROC] across sites 0.93–0.94), 30-day unplanned readmission (AUROC 0.75–0.76), prolonged length of stay (AUROC 0.85–0.86), and all of a patient's final discharge diagnoses (frequency-weighted AUROC 0.90). These models outperformed traditional, clinically-used predictive models in all cases. We believe that this approach can be used to create accurate and scalable predictions for a variety of clinical scenarios. In a case study of a particular prediction, we demonstrate that neural networks can be used to identify relevant information from the patient's chart.

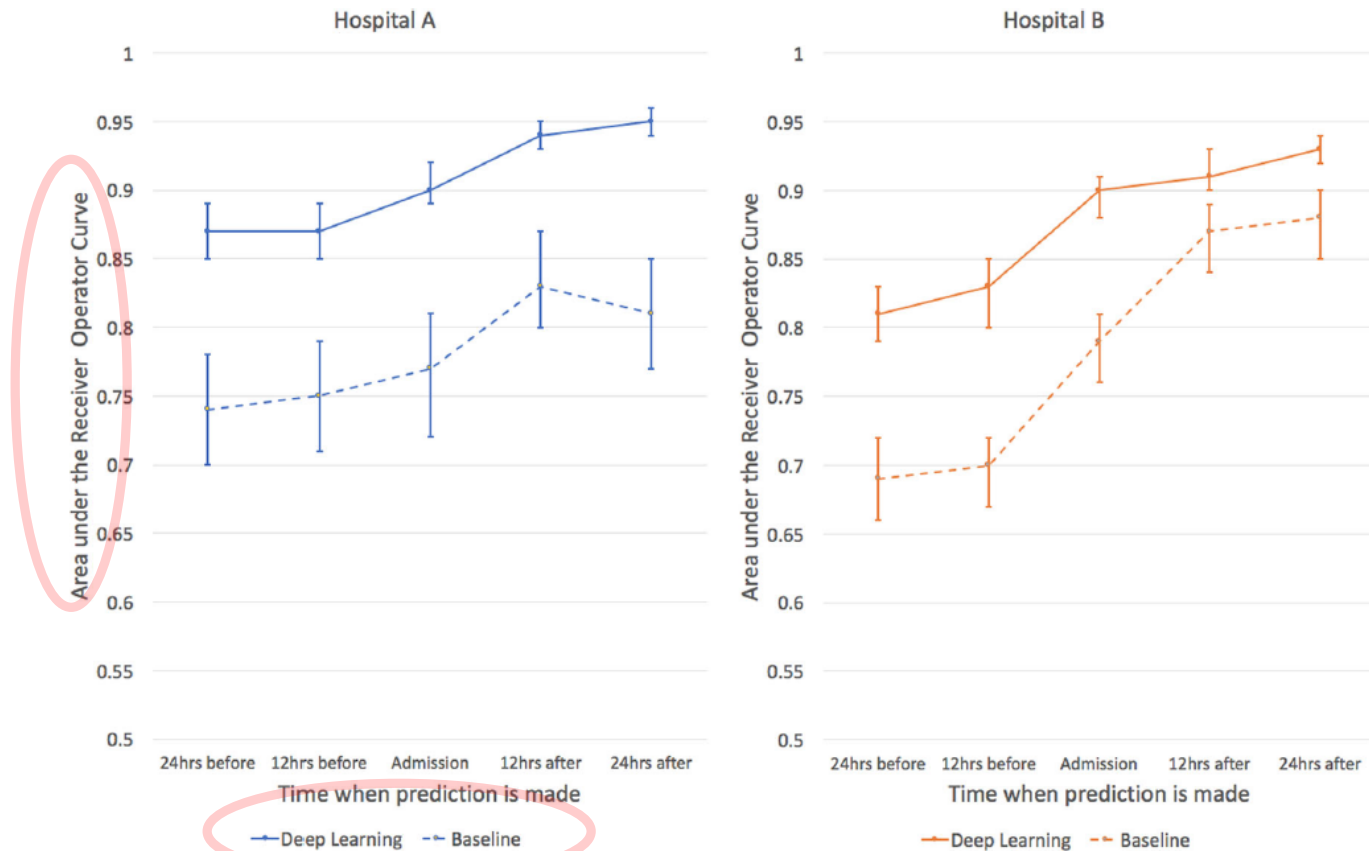


**Fig. 2** The area under the receiver operating characteristic curves are shown for predictions of inpatient mortality made by deep learning and baseline models at 12 h increments before and after hospital admission. For inpatient mortality, the deep learning model achieves higher discrimination at every prediction time compared to the baseline for both the University of California, San Francisco (UCSF) and University of Chicago Medicine (UCM) cohorts. Both models improve in the first 24 h, but the deep learning model achieves a similar level of accuracy approximately 24 h earlier for UCM and even 48 h earlier for UCSF. The error bars represent the bootstrapped 95% confidence interval

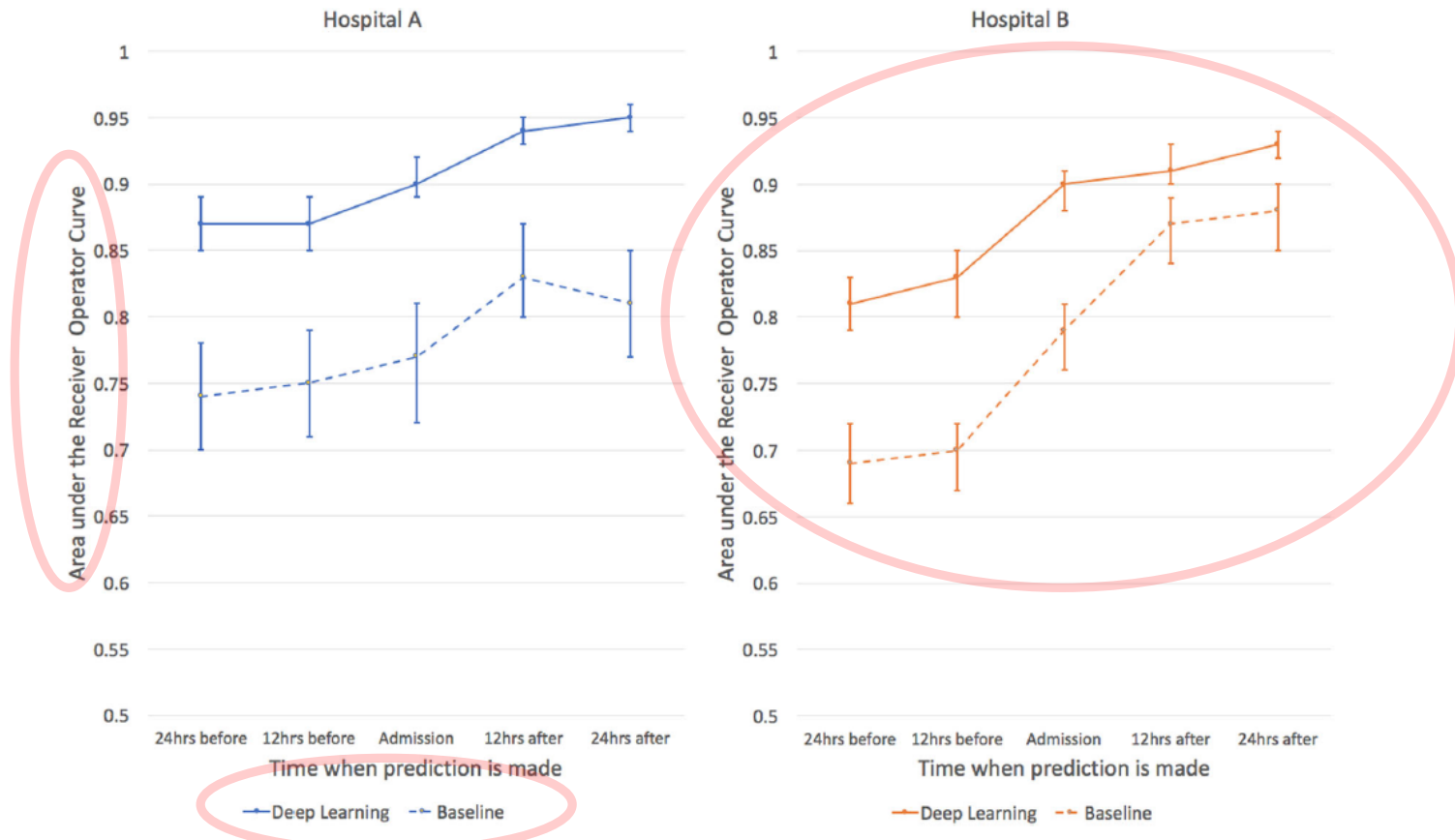


**Fig. 2** The area under the receiver operating characteristic curves are shown for predictions of inpatient mortality made by deep learning and baseline models at 12 h increments before and after hospital admission. For inpatient mortality, the deep learning model achieves higher discrimination at every prediction time compared to the baseline for both the University of California, San Francisco (UCSF) and University of Chicago Medicine (UCM) cohorts. Both models improve in the first 24 h, but the deep learning model achieves a similar level of accuracy approximately 24 h earlier for UCM and even 48 h earlier for UCSF. The error bars represent the bootstrapped 95% confidence interval



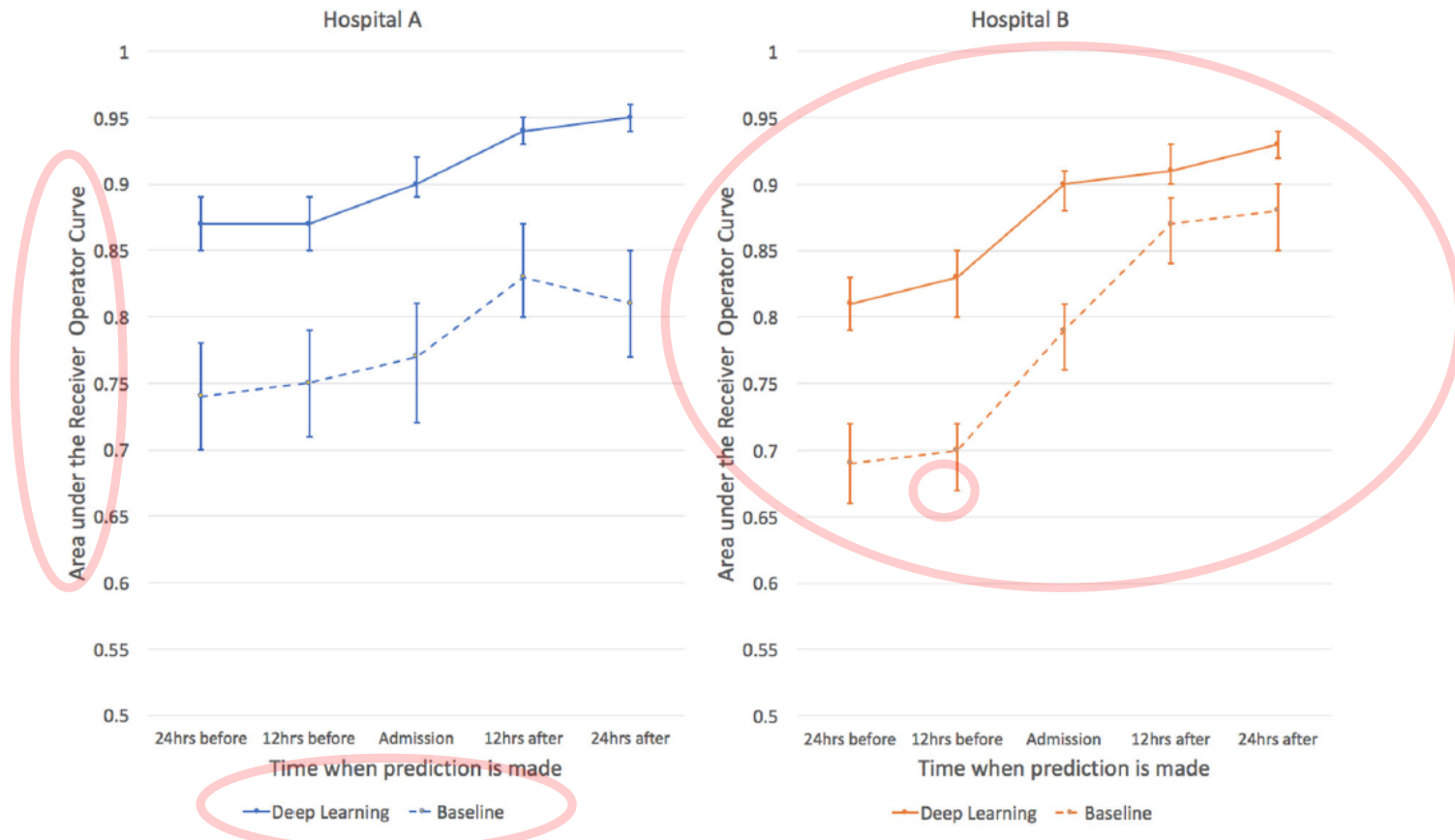


**Fig. 2** The area under the receiver operating characteristic curves are shown for predictions of inpatient mortality made by deep learning and baseline models at 12 h increments before and after hospital admission. For inpatient mortality, the deep learning model achieves higher discrimination at every prediction time compared to the baseline for both the University of California, San Francisco (UCSF) and University of Chicago Medicine (UCM) cohorts. Both models improve in the first 24 h, but the deep learning model achieves a similar level of accuracy approximately 24 h earlier for UCM and even 48 h earlier for UCSF. The error bars represent the bootstrapped 95% confidence interval



**Fig. 2** The area under the receiver operating characteristic curves are shown for predictions of inpatient mortality made by deep learning and baseline models at 12 h increments before and after hospital admission. For inpatient mortality, the deep learning model achieves higher discrimination at every prediction time compared to the baseline for both the University of California, San Francisco (UCSF) and University of Chicago Medicine (UCM) cohorts. Both models improve in the first 24 h, but the deep learning model achieves a similar level of accuracy approximately 24 h earlier for UCM and even 48 h earlier for UCSF. The error bars represent the bootstrapped 95% confidence interval





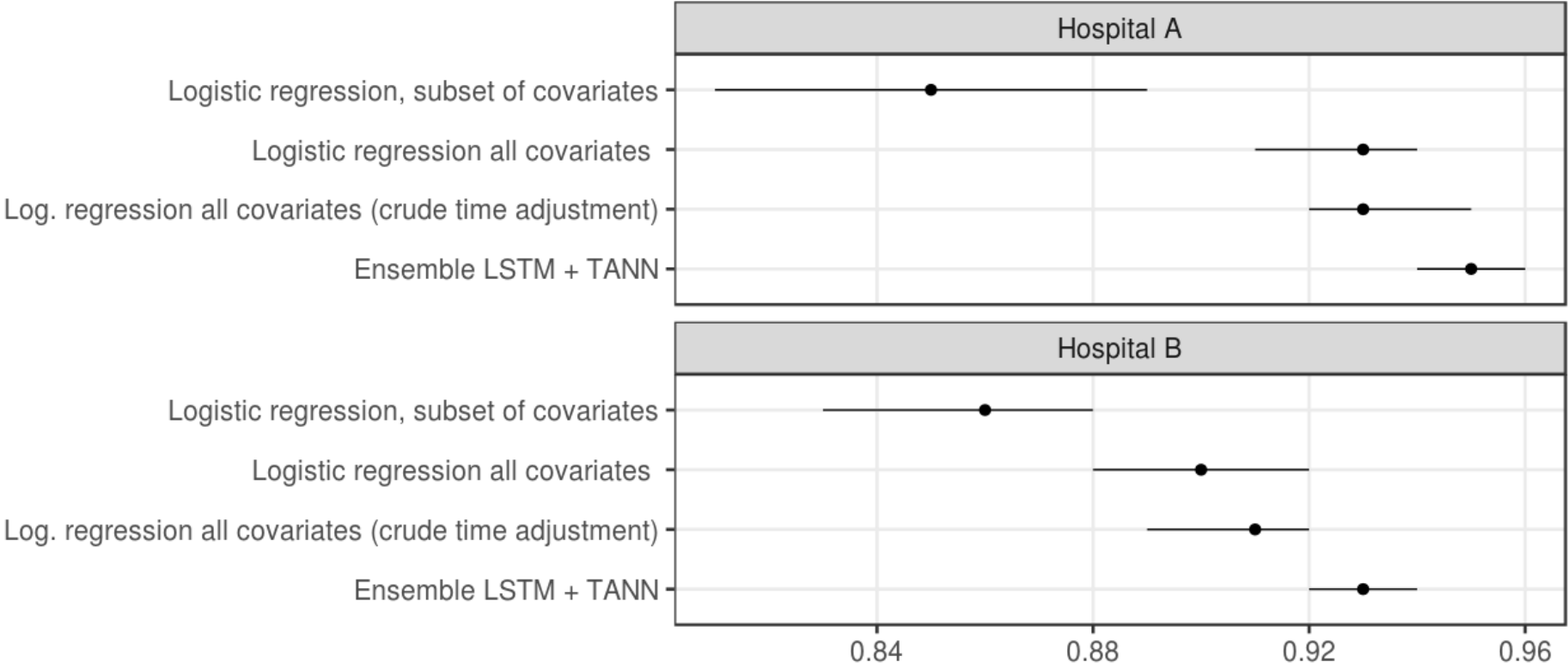
**Fig. 2** The area under the receiver operating characteristic curves are shown for predictions of inpatient mortality made by deep learning and baseline models at 12 h increments before and after hospital admission. For inpatient mortality, the deep learning model achieves higher discrimination at every prediction time compared to the baseline for both the University of California, San Francisco (UCSF) and University of Chicago Medicine (UCM) cohorts. Both models improve in the first 24 h, but the deep learning model achieves a similar level of accuracy approximately 24 h earlier for UCM and even 48 h earlier for UCSF. The error bars represent the bootstrapped 95% confidence interval

Supplemental Table 1: Prediction accuracy of each task of deep learning model compared to baselines

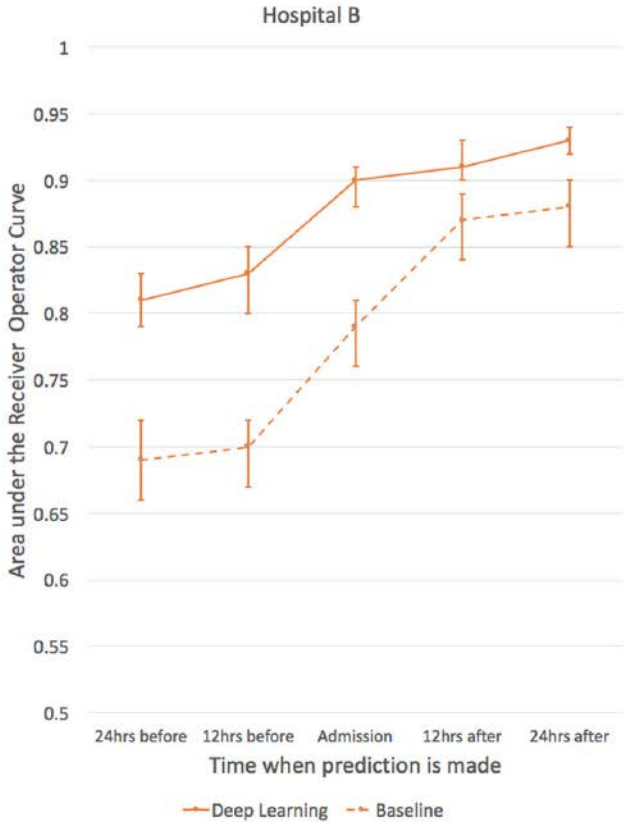
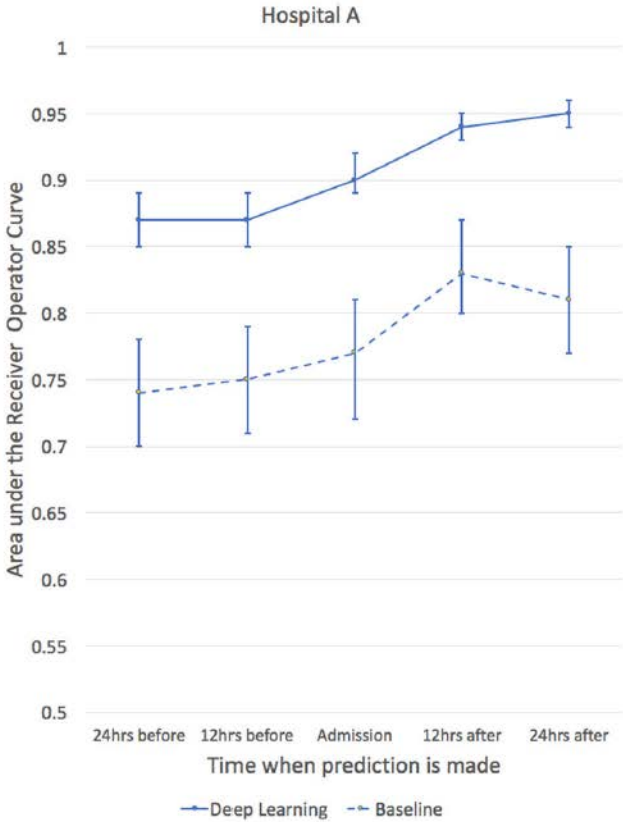
	Hospital A	Hospital B
<b>Inpatient Mortality, AUROC<sup>1</sup>(95% CI)</b>		
Deep learning 24 hours after admission	<b>0.95</b> (0.94-0.96)	<b>0.93</b> (0.92-0.94)
Full feature enhanced baseline at 24 hours after admission	0.93 (0.92-0.95)	0.91 (0.89-0.92)
Full feature simple baseline at 24 hours after admission	0.93 (0.91-0.94)	0.90 (0.88-0.92)
Baseline (aEWS <sup>2</sup> ) at 24 hours after admission	0.85 (0.81-0.89)	0.86 (0.83-0.88)
<b>30-day Readmission, AUROC (95% CI)</b>		
Deep learning at discharge	<b>0.77</b> (0.75-0.78)	<b>0.76</b> (0.75-0.77)
Full feature enhanced baseline at discharge	0.75 (0.73-0.76)	0.75 (0.74-0.76)
Full feature simple baseline at discharge	0.74 (0.73-0.76)	0.73 (0.72-0.74)
Baseline (mHOSPITAL <sup>3</sup> ) at discharge	0.70 (0.68-0.72)	0.68 (0.67-0.69)
<b>Length of Stay at least 7 days AUROC (95% CI)</b>		
Deep learning 24 hours after admission	<b>0.86</b> (0.86-0.87)	<b>0.85</b> (0.85-0.86)
Full feature enhanced baseline at 24 hours after admission	0.85 (0.84-0.85)	0.83 (0.83-0.84)
Full feature simple baseline at 24 hours after admission	0.83 (0.82-0.84)	0.81 (0.80-0.82)
Baseline (mLiu <sup>4</sup> ) at 24 hours after admission	0.76 (0.75-0.77)	0.74 (0.73-0.75)

<sup>1</sup> Area under the receiver operator curve<sup>2</sup> Augmented early warning score<sup>3</sup> Modified HOSPITAL score<sup>4</sup> Modified Liu score

# Optimised deep learning ensembles are not the driver of "accuracy"



# We presented our “best” model and compared it to an inappropriate baseline

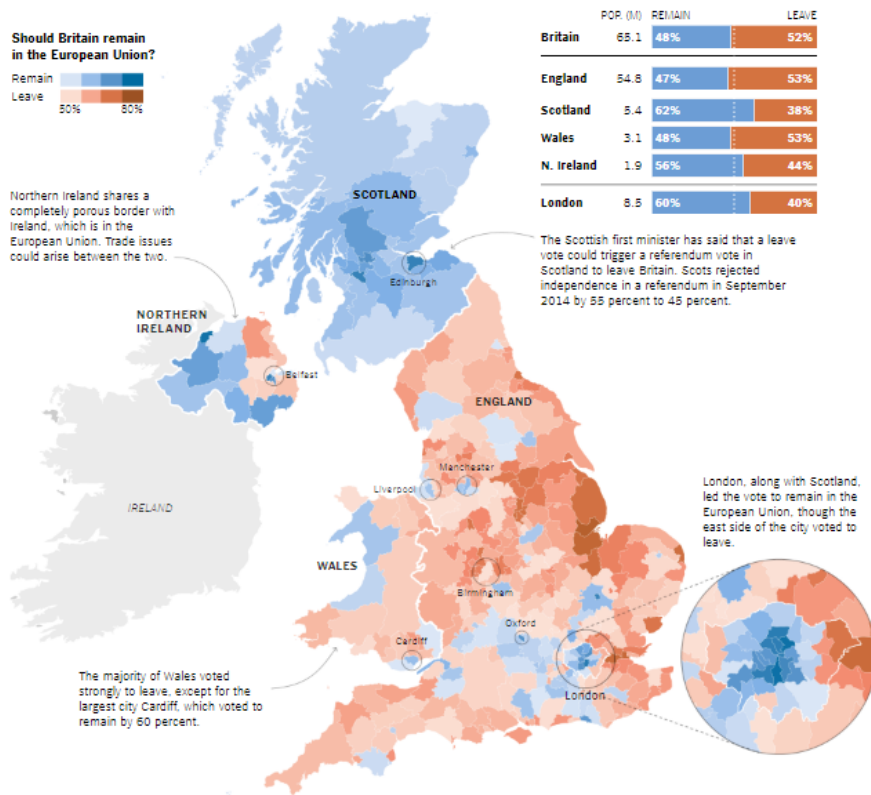


# Effective visualisation is important

## How Britain Voted in the E.U. Referendum

By GREGOR AISCH, ADAM PEARCE and KARL RUSSELL. UPDATED June 24, 2016

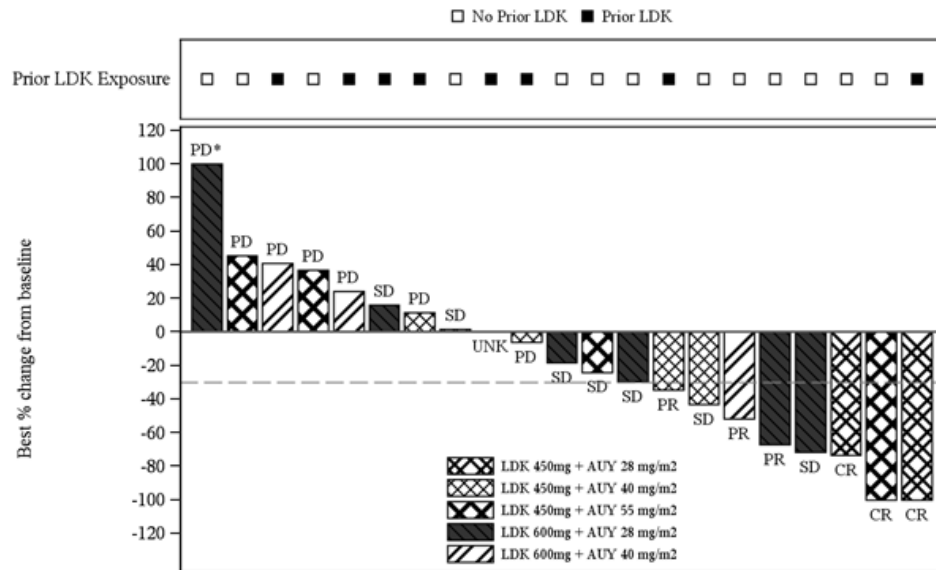
Britons voted on Thursday to leave the European Union. The Leave side led with 17.4 million votes, or 52 percent, versus the Remain side's 16.1 million, or 48 percent, with a turnout of around 72 percent. [RELATED ARTICLE](#)



<https://www.nytimes.com/interactive/2016/06/24/world/europe/how-britain-voted-brexit-referendum.html>

# We are not always good at it

Figure 11-1 (Page 1 of 1)  
 Best percentage change from baseline in sum of longest diameters and best overall response as per investigator by prior LDK378 treatment (Full analysis set)



\* Denotes the percentage change from baseline greater than 100.

Source: Table 11-4, Listing 14.2-1.2 and Listing 16.2.4-1.5



# Visualisation panel

“Visualization and the use of graphics can help at every stage of an analysis, from the planning and design of an experiment, the very first data explorations, through to the communication of conclusions and recommendations.

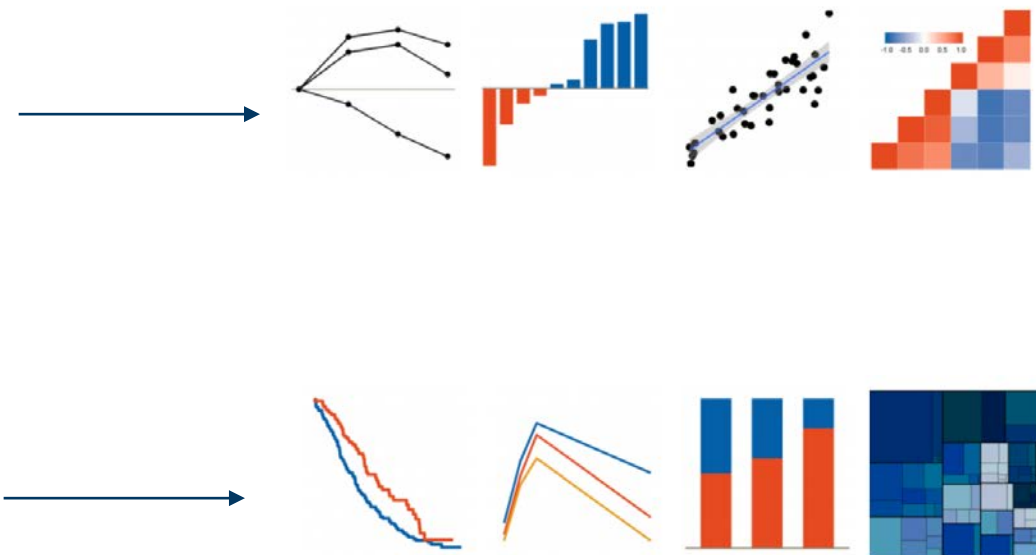
Visualization is more than "plotting data"; it can lead to a deeper understanding and inform next steps.

The role of the STRATOS visualization panel is to promote the use of good graphical principles for effective visual communication, providing guidance and recommendations covering all aspects from the design, implementation and review of statistical graphics.”

# Effective visualisation is important throughout the workflow

## Topic groups

1	Missing data
2	Selection of variables and functional forms in multivariable analysis
3	Initial data analysis
4	Measurement error and misclassification
5	Study design
6	Evaluating diagnostic tests and prediction models
7	Causal inference
8	Survival analysis
9	High-dimensional data

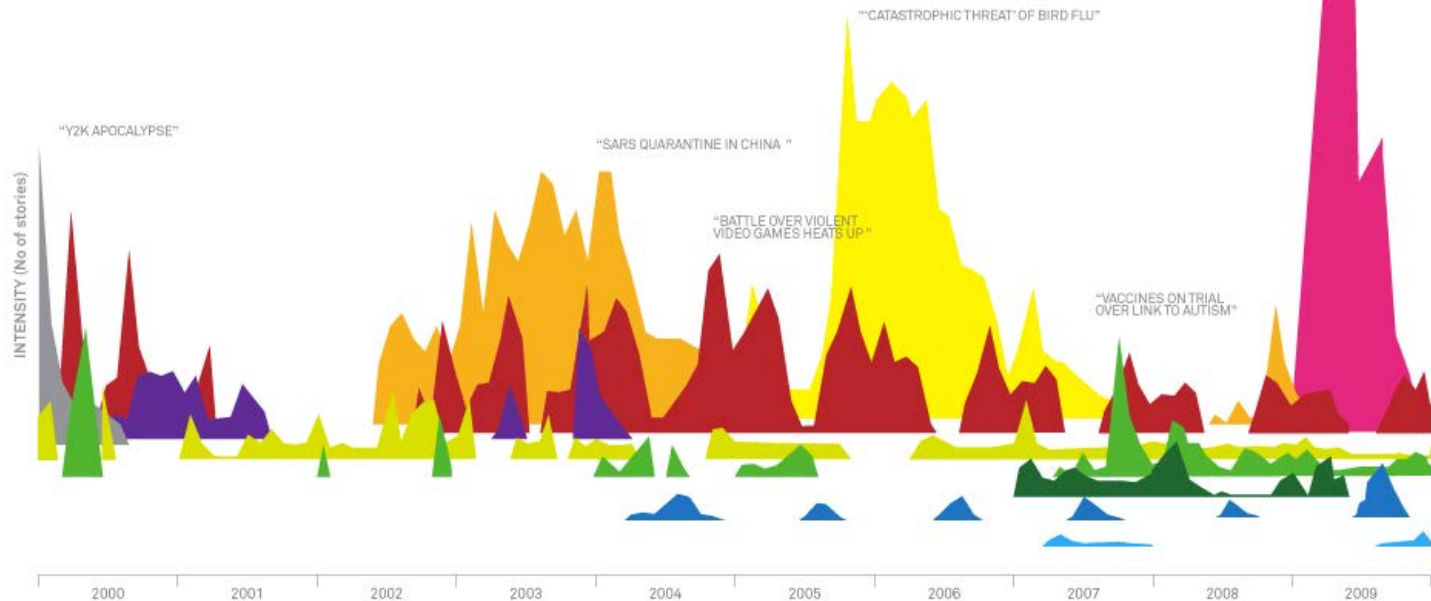


# Elements of the initiative

- Graphical principles and thinking
  1. Graphics Principles Cheat Sheet
  2. Newsletter
- Easing the implementation
  3. Graph Gallery
  4. Analysis Results Datasets
  5. Standardization of most common/important graphs
- Graphics tomorrow ... or today?
  6. Question-based visualizations and interactive graphics
- ...plus overarching stakeholder management and communication

# Beautiful but effective?

"BRITAIN PREPARES FOR 85000 DEATHS FROM SWINE FLU"

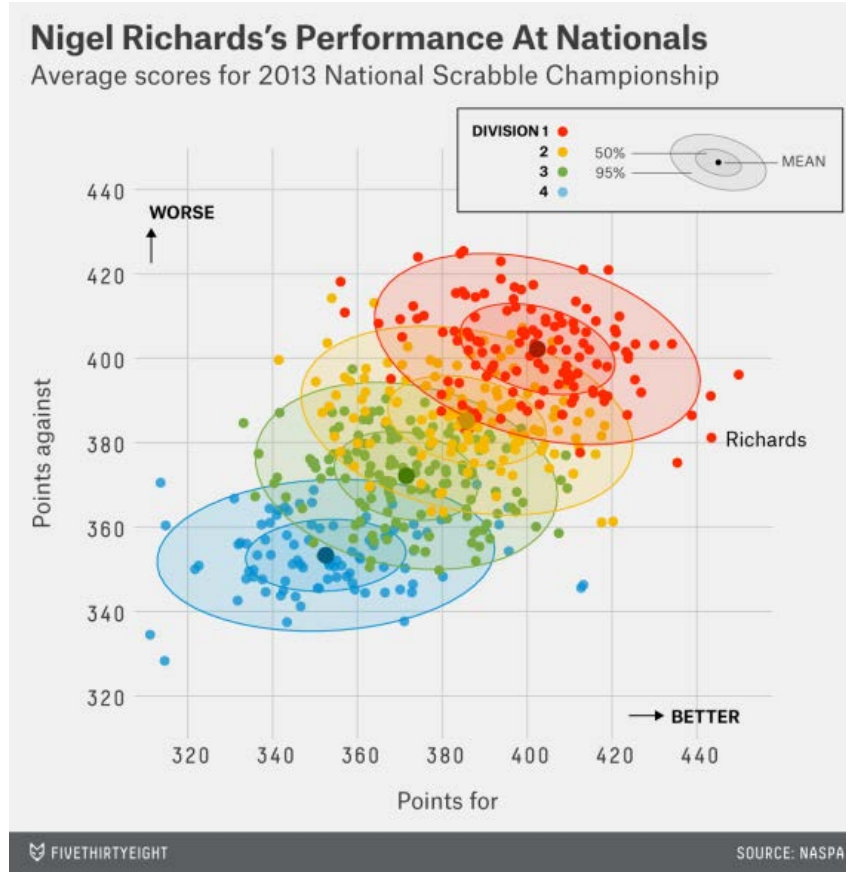


## Story (worldwide deaths)



**Mountains Out of Molehills**  
A timeline of global media scare stories.

# Beautiful and effective?



# Effective data visualisation is effective visual communication

- Effective graphs...
  - are visually appealing, intuitive, legible
  - use the correct graph type and axis scales
  - use proximity & alignment to facilitate comparison
  - use labels and annotations to add clarity to the message
- Most importantly, effective use of visualisations
  - Enables clear and impactful communication
  - Elevates influence with stakeholders
  - Facilitates informed decision making



# Principles for effective visual communication

## Graphical Principles Cheat Sheet

Authors: Mark Bailille,<sup>1</sup> Alison Margolske,<sup>1</sup> Baldur Magnusson,<sup>1</sup> Andrew Wright,<sup>1</sup> Ruquan You,<sup>2</sup> Ivan-Toma Vranesic,<sup>1</sup> Marc Vandemeulebroecke<sup>1</sup>  
Affiliations: <sup>1</sup>Novartis Pharma AG, Basel, Switzerland; <sup>2</sup>Novartis Institutes for Biomedical Research, Cambridge, MA, United States; <sup>3</sup>Novartis Institutes for Biomedical Research, Shanghai, China

### 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 cluttering. 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

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

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

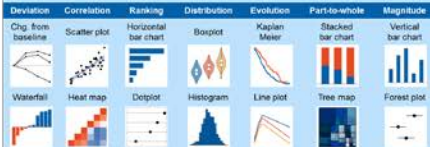
**Familiarity** – using familiar styles, icons, navigation structure makes viewers feel confident

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

### 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



Utilize the **Graph Gallery** ([ourstory.novartis.com](https://www.novartis.com/our-story/innovation/our-people/our-principles)). There are many effective base designs that can be adapted to your purpose.

### Implementation Considerations

Plot axes 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. data 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, CI).

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 discontinuity 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. Plot dates or precise duration (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.

### Good graph checklist

- | Clear Communication   | Implementation Considerations   |
|---|---|
| <input type="checkbox"/> Is the message of the graph as clear as possible?  | <input type="checkbox"/> Are multiple panels plotted on the same scale?   |
| <input type="checkbox"/> Is it easy for someone unfamiliar with the data to interpret the graph?                  | <input type="checkbox"/> Are logarithmically distributed variables plotted on a log scale?                                  |
| <input type="checkbox"/> Are the patterns/relationships easily identified?  | <input type="checkbox"/> Are common baselines used wherever possible?   |
| <input type="checkbox"/> Is the graph tailored to its primary purpose and audience?                               | <input type="checkbox"/> Does the orientation of the axes aid interpretation?   |
| <input type="checkbox"/> Is the correct graph type used?  | <input type="checkbox"/> Does the aspect ratio allow the reader to see variations in the data?                              |
|   | <input type="checkbox"/> Are data across a disconnected time scale clearly disconnected?                                    |
|   | <input type="checkbox"/> Are data spaced proportionally to the actual time interval (instead of according to visit number)? |
|   | <input type="checkbox"/> Are data and inferences plotted to support stories about models?                                   |
|   | <input type="checkbox"/> Are a number of patients by group reported if this adds context?                                   |
| Facilitating Comparisons  | Legibility and Clarity  |
| <input type="checkbox"/> Are elements to be compared grouped together?  | <input type="checkbox"/> Can all graphical elements be seen?  |
| <input type="checkbox"/> Are labels placed next to data instead of in legends?                                    | <input type="checkbox"/> Does the graph have a clear title, axis labels, annotations and data units?                        |
| <input type="checkbox"/> Have categories been ordered for easy comparison?  | <input type="checkbox"/> Can the font be read without eye strain or effort?   |
| <input type="checkbox"/> Can the plot be read without doing mental calculations?                                  | <input type="checkbox"/> Are sans-serif fonts used?   |
| <input type="checkbox"/> Are the estimates of interest plotted (e.g. mean differences with confidence intervals)? | <input type="checkbox"/> Do text sizes have correct hierarchy (big to small, main text to subtitle)?                        |
| Color for emphasis or distinction   | <input type="checkbox"/> Are the elements of the graph clearly labeled (e.g. points, error bars, lines, shaded regions)?    |
| <input type="checkbox"/> Are graphical elements displayed in a dark color on a light background?                  | <input type="checkbox"/> Are labels oriented horizontally where possible?   |
| <input type="checkbox"/> Are grid lines drawn with a thin line and a light color such as grey?                    |   |
| <input type="checkbox"/> Are colors used sparingly (e.g. max 3)?  |   |
| <input type="checkbox"/> Do all elements in the graph have a purpose (e.g. colors, textures, grid lines)?         |   |
| <input type="checkbox"/> Are the same colors of used to mean the same thing in a series of graphs?                |   |

### Facilitating Comparisons

#### Proximity improves association

Place labels next to data instead of using legends.

Group together elements to be compared directly.

#### 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.

### 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.



Least accurate → Most accurate

volume charts, poorly designed heat maps, multivariate density plots, bubble charts, line graphs, bar charts, pie charts, mosaic charts, stacked bar charts, small coordinate charts, dot plots, bar charts, multiple plots

### 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 distracting. Use white background and try using other methods to distinguish different curves.

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.

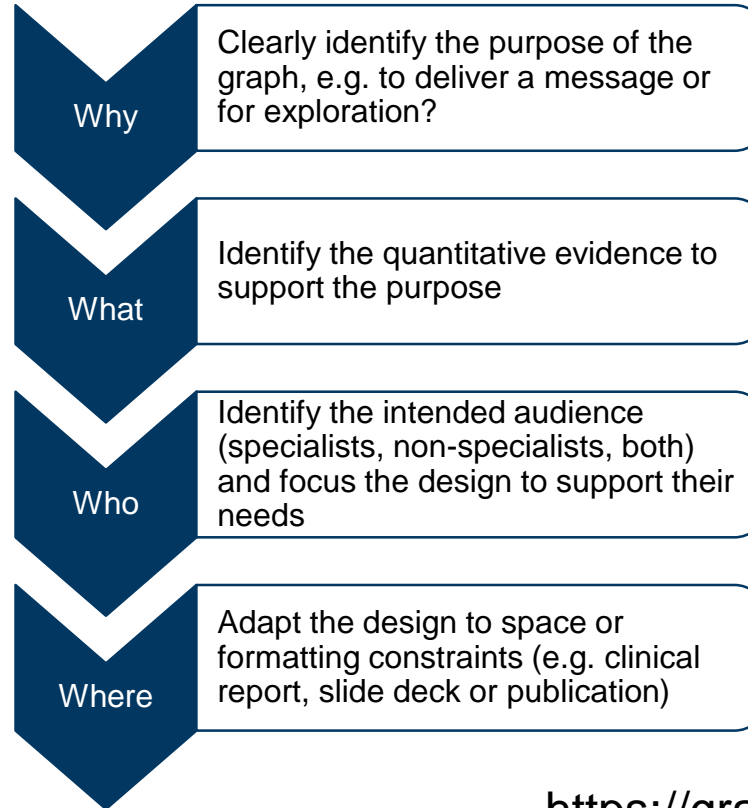
One solution could be repeating the data in a darker color.

### 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 applications to the development of graphical methods*, USA, Vol. 79, No. 377-39, 537–554, 1984.  
S. Few, *Show Me The Numbers - Designing Tables and Graphs to Enlighten* (2nd Edition), Berkeley, CA: Analytics Press, 2012.  
M. Wang, *The Wall Street Journal Guide to Information Graphics: The One and Only of Its-Kind Data, Facts, and Figures*, December 16, 2013.  
J. Deonant, *Tables, Maps, and Graphs: Effective communication for reform minds*, PRINCIPALIE.  
N. B. Robbins, *Creating More Effective Graphs*, Chart House.

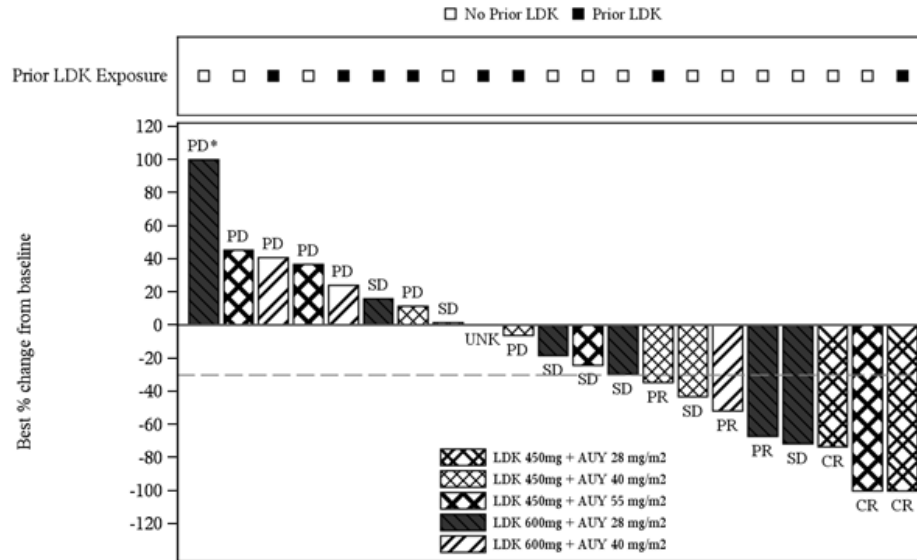
Online resources:  
<https://www.dccproc.com/> (S. Few) | <https://www.analyticspress.com/> (N. Yau) | <http://www.theinformationmarket.com/> (A. Cavali)  
<https://www.strobilini.com/en/it/> (S. Tufte) | <https://www.chartwell.com/> (A. Deonant)

# Use the cheat sheet for design and planning



# Use the cheat sheet for critical review

Figure 11-1 (Page 1 of 1)  
 Best percentage change from baseline in sum of longest diameters and best overall response as per investigator by prior LDK378 treatment (Full analysis set)



- \* Denotes the percentage change from baseline greater than 100.  
 Source: Table 11-4, Listing 14.2.1.2 and Listing 16.2.4.1.5

### 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

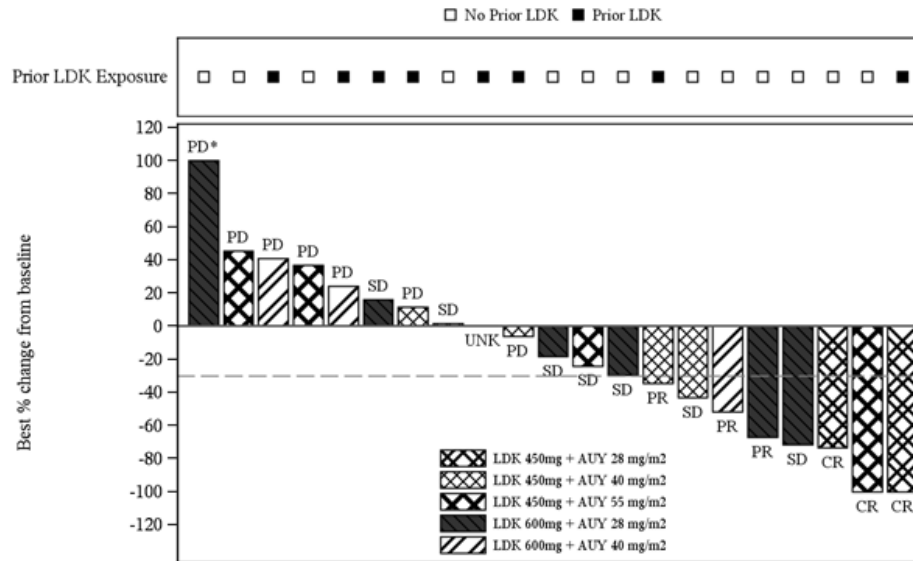
**Reduce mental arithmetic**

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

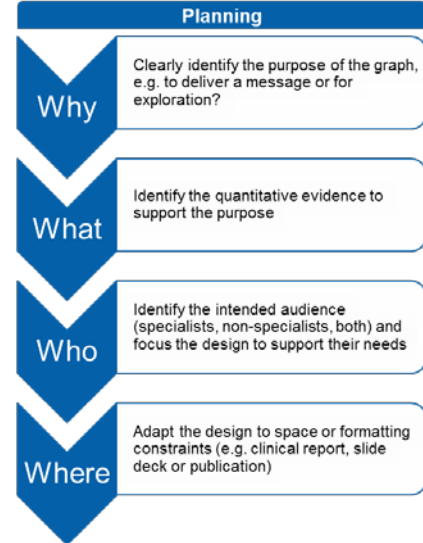
Use reference lines and other visual anchors.

# Use the cheat sheet for critical review

Figure 11-1 (Page 1 of 1)  
 Best percentage change from baseline in sum of longest diameters and best overall response  
 as per investigator by prior LDK378 treatment  
 (Full analysis set)



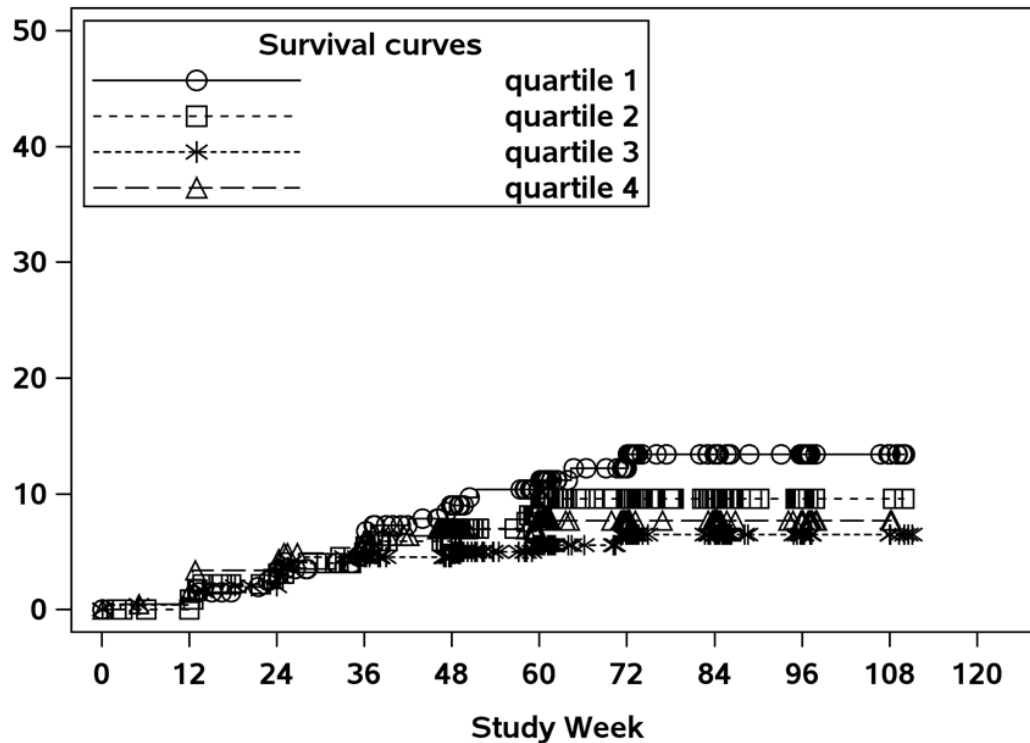
- \* Denotes the percentage change from baseline greater than 100.  
 Source: Table 11-4, Listing 14.2.1.2 and Listing 16.2.4.1.5



# This is a continual process

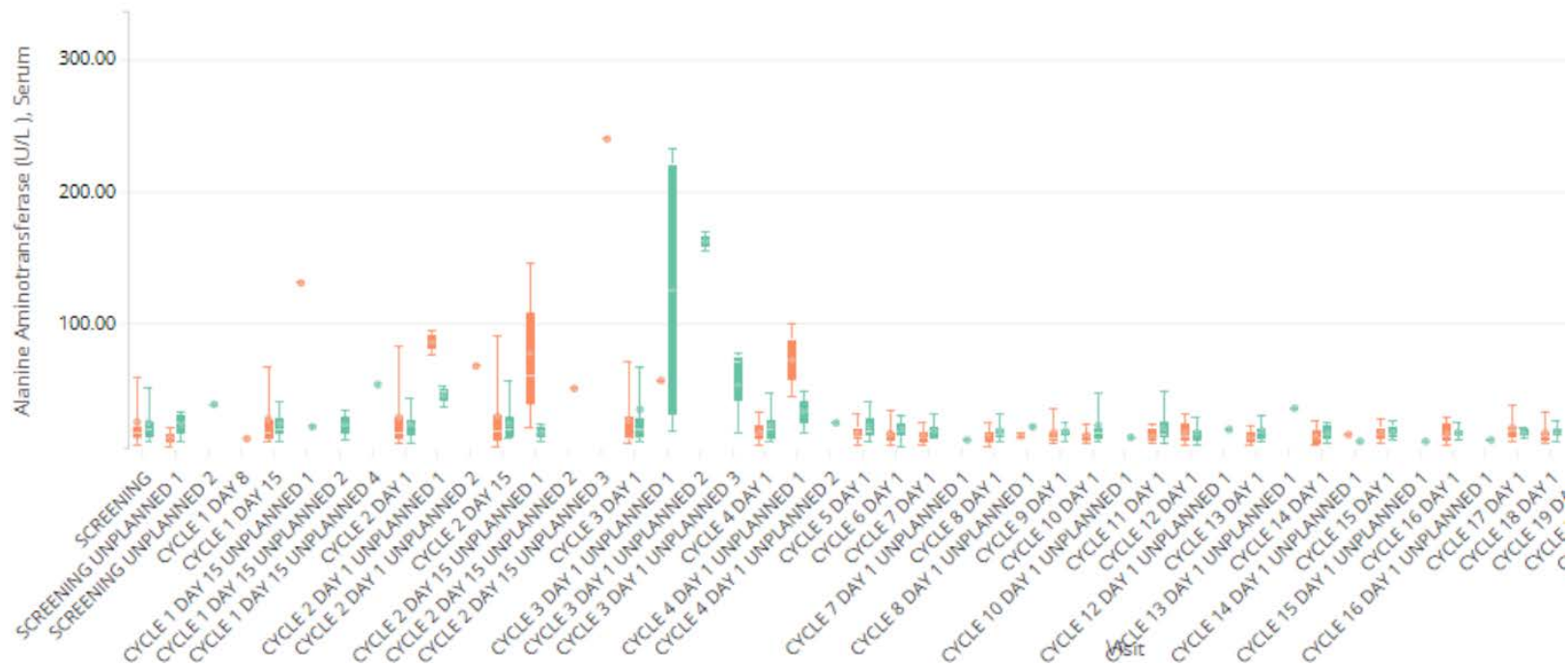
Planned Treatment:

mg



# This is a continual process

50 of 50 participant(s) shown (100.0%)





# Three laws for improving visual communication

## Have a clear purpose

- Know the purpose of creating the graph
- Identify the quantitative evidence to support the purpose
- Identify the audience and focus the design to support their needs

## Show the data clearly

- Choose the appropriate graph type to display your data
- Avoid misrepresentation (use appropriate scales)
- Maximize data to ink ratio (reduce distraction, less is more)

## Make the message obvious

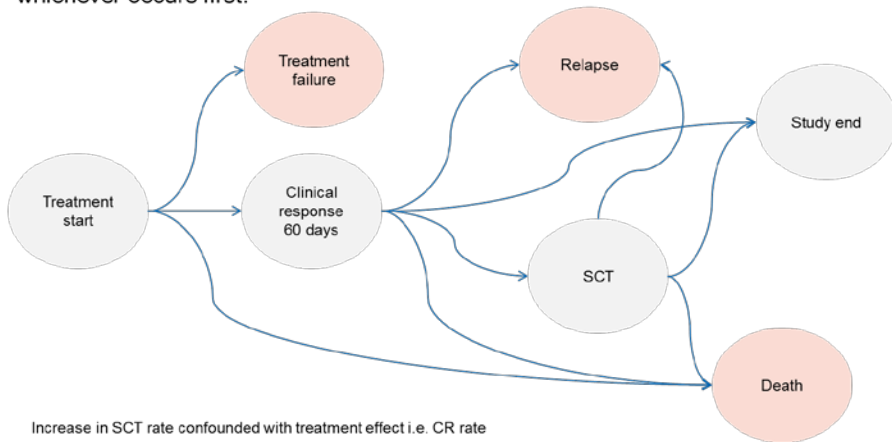
- Use proximity and alignment to aid in comparisons
- Minimize mental arithmetic (e.g. plot the difference)
- Use colors and annotations to highlight important details

<https://arxiv.org/abs/1903.09512>

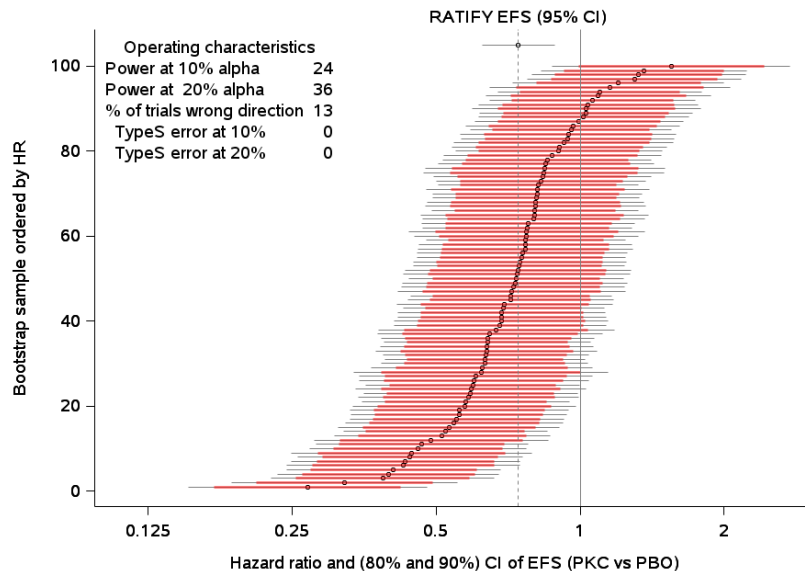
# Effective visualization is important for planning

## Event free survival endpoint

An EFS event for the key secondary endpoint is defined as a failure to achieve a CR within 60 days of study treatment, relapse from CR, or death due to any cause, whichever occurs first.



60 sampled RATIFY patients (1:1 ratio) with poor FLT3 imbalance



# Effective visualization is important during exploratory analysis

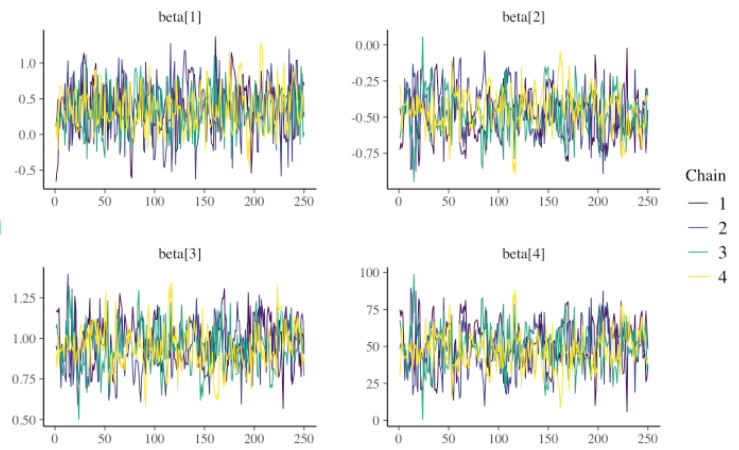


Original Article | [Free Access](#)

## Visualization in Bayesian workflow

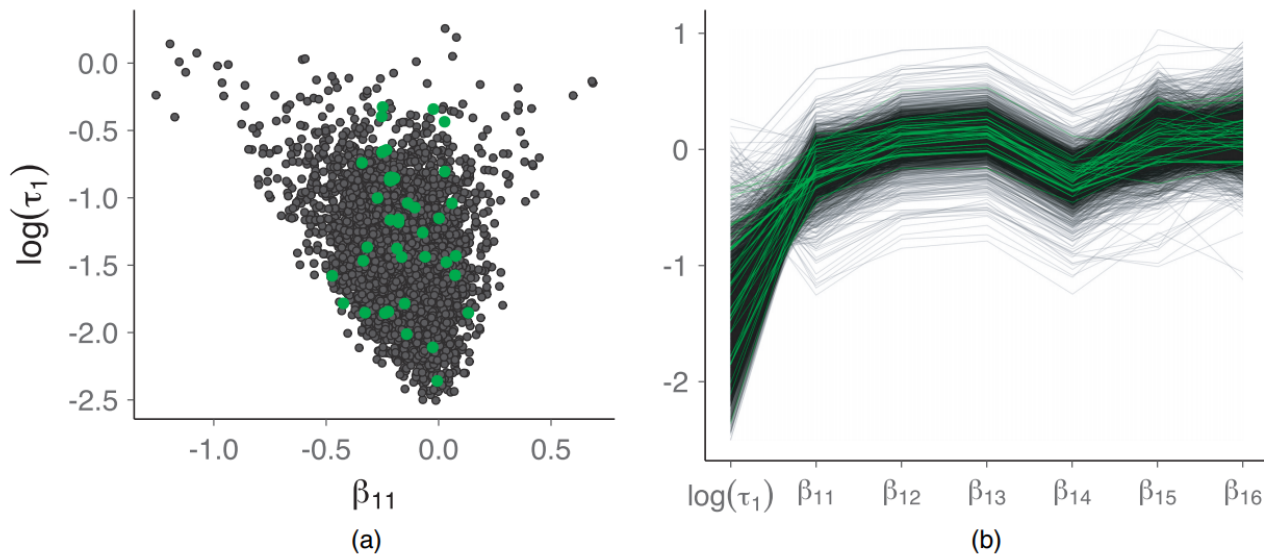
Jonah Gabry , Daniel Simpson, Aki Vehtari, Michael Betancourt, Andrew Gelman

First published: 15 January 2019 | <https://doi.org/10.1111/rssa.12378>



<https://mc-stan.org/bayesplot/reference/MCMC-traces.html>

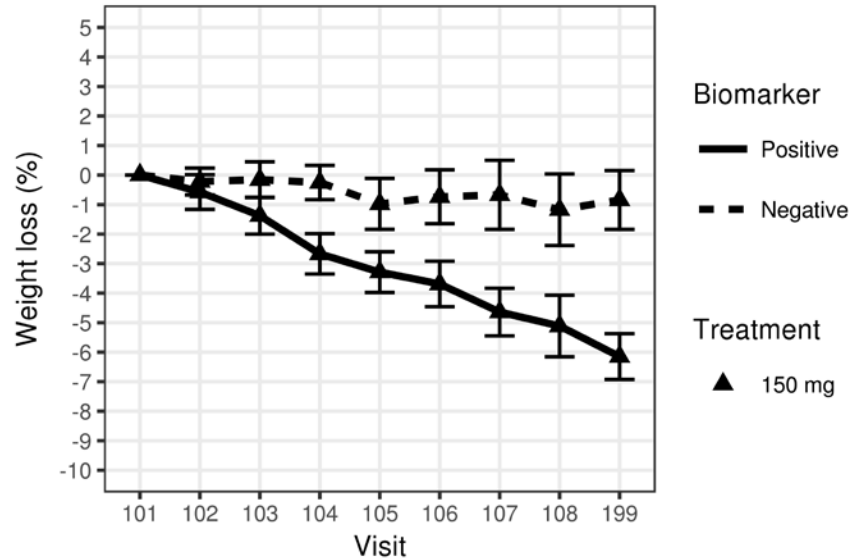
# Effective visualization is important during exploratory analysis



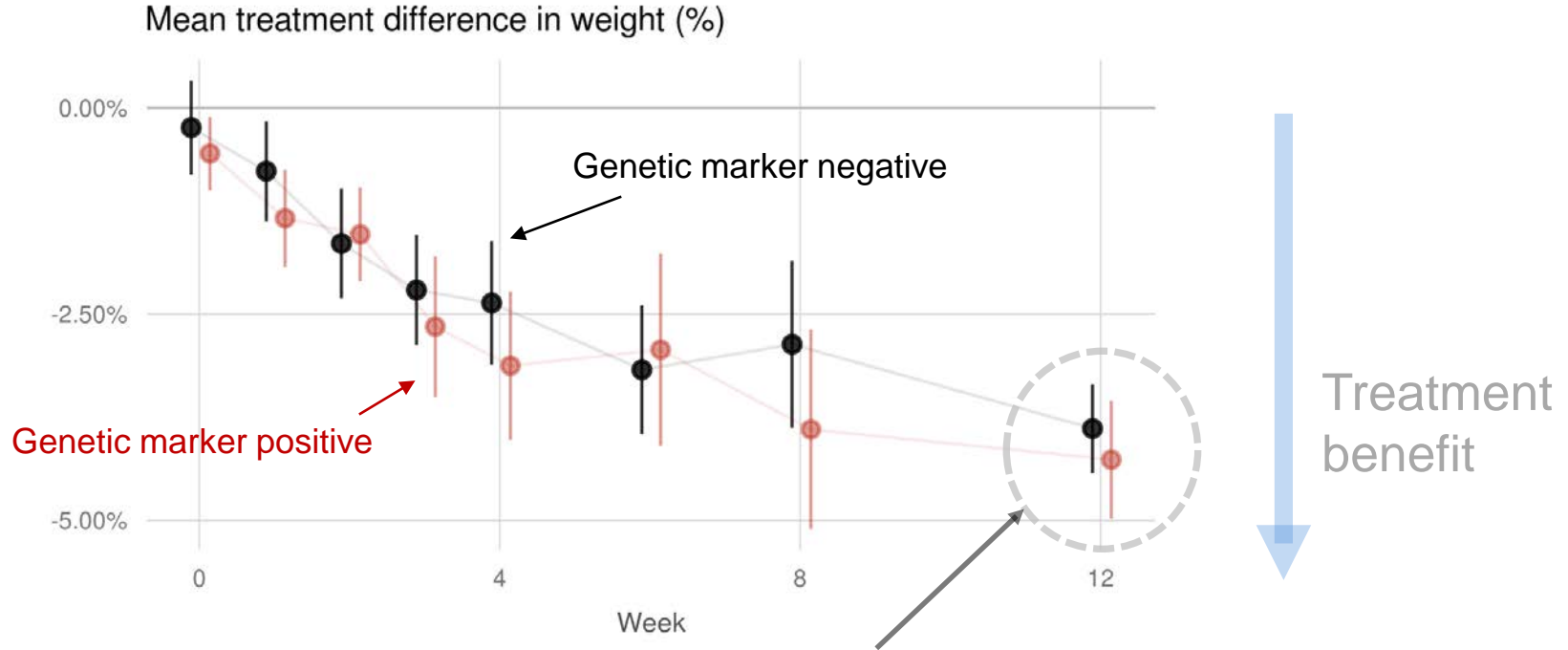
**Fig. 5.** Diagnostic plots for HMC sampling (models were fitted by using the RStan interface to Stan 2.17 (Stan Development Team, 2017a)): (a) for model 3, a bivariate plot of the log-standard-deviation of the cluster level slopes ( $y$ -axis) against the slope for the first cluster ( $x$ -axis) (the green dots indicate starting points of divergent transitions; this plot can be made by using `mcmc_scatter` in `bayesplot`); (b) for model 3, a parallel co-ordinates plot showing the cluster level slope parameters and their log-standard-deviation  $\log(\tau_1)$  (the green lines indicate starting points of divergent transitions; this plot can be made by using `mcmc_parcoord` in `bayesplot`)

# Effective visualization important for reporting

%improvement in baseline weight through week 12 by subgroup



# Genetic marker positive is **not** predictive of treatment response



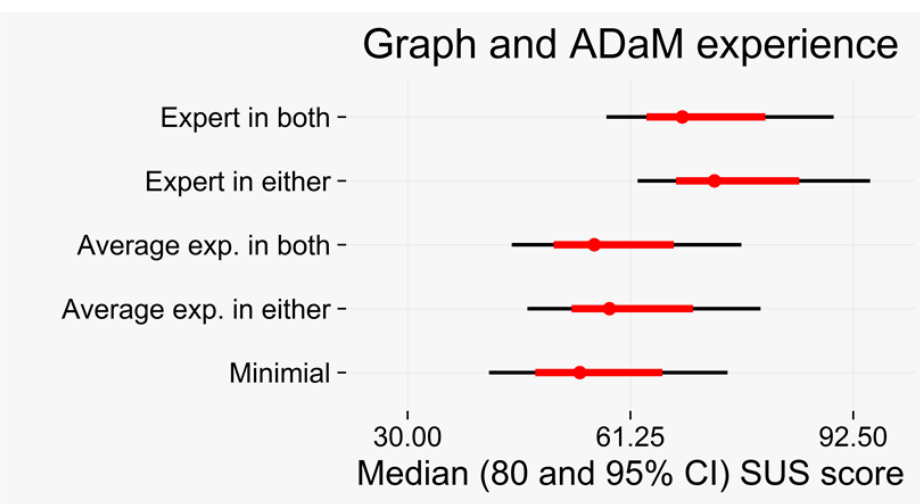
The average treatment effect is **similar** in **both** the genetic marker positive and negative subgroups and does **not** warrant further investigation

# How can the VP help across skill levels?

A survey was sent to associates working with clinical data

The purpose to:

- evaluate ADaM as a data standard for graph production
- Identify key issues associates currently experience
- Explore issues related to role and experience level
- 85 respondents



# Provide access to examples with code and data



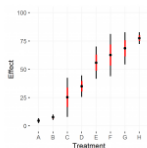
Email  Password

[Register](#), if you have not already; [Reset password](#), if you forgot it

[Home](#) [Samples](#) [Gallery](#) [Discussion Forum](#) [Upload Request](#) [Request for Help](#) [Links](#) [FAQ](#)

## 127 Results

Order by: [Views](#) | [Downloads](#) - [Rating](#)

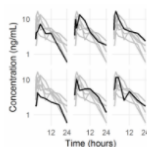


### 1593: Revising a dynamite plot

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

[R Version 3.2.3](#) • [Barchart](#) • [Cheatsheet](#) • [Continuous](#) • [Dotplot](#) • [Efficacy](#) • [Forestplot](#) • [GGPLOT2](#) • [Model-based](#) • [R](#) • [Safety](#)

Views: 108



### 1586: Demonstration of a multi panel (small multiples) line plot

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

[R Version 3.2.3](#) • [Cheatsheet](#) • [Continuous](#) • [GGPLOT2](#) • [Lineplot](#) • [Model-based](#) • [Panel](#) • [R](#) • [Time Line](#)

Views: 100

### Search

### Filter

#### Logic

AND  OR

#### Language

R  SAS  
 MATLAB  all

#### Data Types

Binary  Continuous  
 Mixed  Model-based  
 Ordinal  Time-To-Event

#### Sample Type

Barchart  Boxplot  
 Distribution  Dotplot  
 Forestplot  Lineplot  
 Other  Panel  
 Patient profile/listing  Scatterplot  
 Survival  Spaghetti  
 Time Line



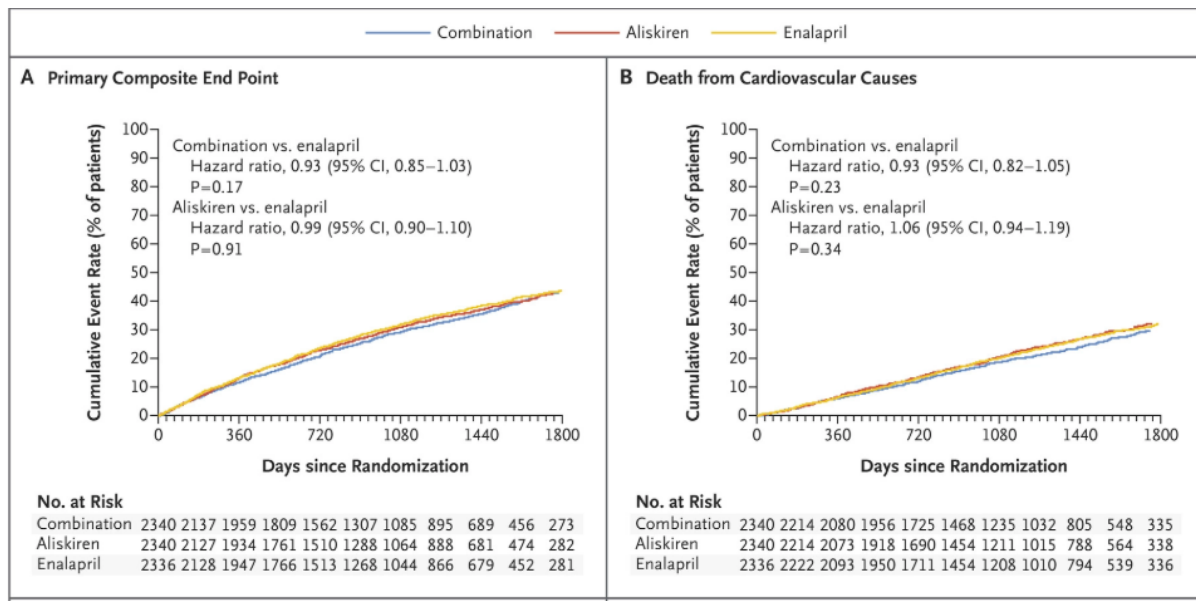
# How can the VP help with implementation across skill levels?

(Negative) qualitative comments

- ***Complex graphs such as Forest Plots will need a fair amount of data manipulation to get all needed for input to the graph***
- ***some variables needed for graphs are not in ADaM datasets***
- ***Trying to figure out what the different parameters mean and **extracting the information relevant to my task.*****

# Aliskiren, Enalapril, or Aliskiren and Enalapril in Heart Failure

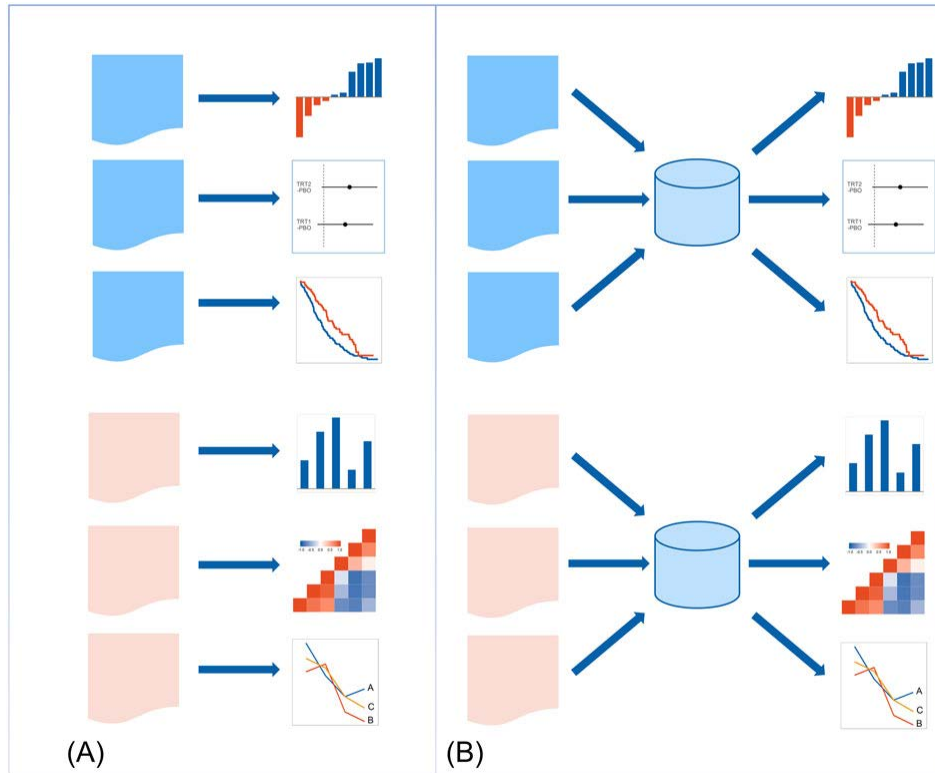
John J.V. McMurray, M.D., Henry Krum, M.B., B.S., Ph.D., William T. Abraham, M.D., Kenneth Dickstein, M.D., Ph.D., Lars V. Køber, M.D., D.M.Sc., Akshay S. Desai, M.D., M.P.H., Scott D. Solomon, M.D., Nicola Greenlaw, M.Sc., M. Atif Ali, B.A., Yanntong Chiang, Ph.D., Qing Shao, Ph.D., Georgia Tarnesby, M.B., B.Chir., [et al.](#), for the ATMOSPHERE Committees Investigators<sup>†</sup>



**Table 2. Protocol-Specified Primary and Secondary Outcomes.\***

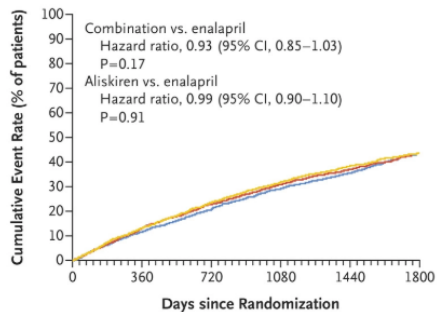
Outcome	Combination Therapy (N = 2340)	Aliskiren (N = 2340)	Enalapril (N = 2336)	Combination Therapy vs. Enalapril		Aliskiren vs. Enalapril	
				Hazard Ratio or Difference (95% CI)	P Value	Hazard Ratio or Difference (95% CI)	P Value
Primary composite outcome: death from cardiovascular causes or first hospitalization for worsening heart failure — no. (%)	770 (32.9)	791 (33.8)	808 (34.6)	0.93 (0.85 to 1.03)	0.17	0.99 (0.90 to 1.10)	0.91
Death from cardiovascular causes	512 (21.9)	562 (24.0)	547 (23.4)	0.93 (0.82 to 1.05)	0.23	1.06 (0.94 to 1.19)	0.34
First hospitalization for worsening heart failure	430 (18.4)	442 (18.9)	452 (19.3)	0.93 (0.82 to 1.06)	0.29	0.99 (0.87 to 1.13)	0.91
Secondary outcome: change in KCCQ clinical summary score at 12 mo <sup>†</sup>	-5.04±0.56	-6.03±0.57	-5.01±0.55	-0.03 (-1.56 to 1.50)	0.97	-1.02 (-2.56 to 0.52)	0.20
Other prespecified exploratory outcomes — no. (%) <sup>‡</sup>							
Death from cardiovascular causes, hospitalization for heart failure, nonfatal myocardial infarction, nonfatal stroke, or resuscitated cardiac arrest	841 (35.9)	874 (37.4)	877 (37.5)	0.94 (0.86 to 1.04)	0.23	1.01 (0.92 to 1.11)	0.80
Fatal or nonfatal myocardial infarction	88 (3.8)	84 (3.6)	100 (4.3)	0.87 (0.66 to 1.16)	0.36	0.85 (0.64 to 1.14)	0.28
Fatal or nonfatal stroke	87 (3.7)	103 (4.4)	93 (4.0)	0.93 (0.70 to 1.25)	0.65	1.12 (0.85 to 1.49)	0.42
First resuscitated cardiac arrest	31 (1.3)	35 (1.5)	32 (1.4)	0.96 (0.58 to 1.57)	0.86	1.10 (0.68 to 1.78)	0.69
Death from any cause	595 (25.4)	654 (27.9)	646 (27.7)	0.91 (0.82 to 1.02)	0.12	1.04 (0.93 to 1.16)	0.46
Composite renal outcome — no. (%) <sup>§</sup>	39 (1.7)	26 (1.1)	18 (0.8)	2.17 (1.24 to 3.79)	0.007	1.50 (0.82 to 2.74)	0.18

# Analysis results data sets



— Combination — Aliskiren — Enalapril

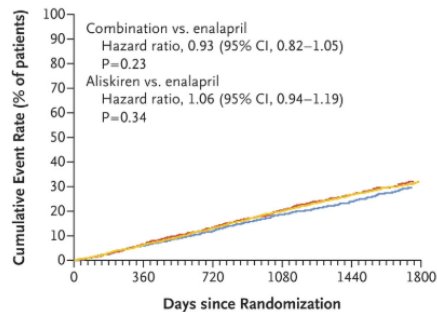
### A Primary Composite End Point



#### No. at Risk

Combination	2340	2137	1959	1809	1562	1307	1085	895	689	456	273
Aliskiren	2340	2127	1934	1761	1510	1288	1064	888	681	474	282
Enalapril	2336	2128	1947	1766	1513	1268	1044	866	679	452	281

### B Death from Cardiovascular Causes



#### No. at Risk

Combination	2340	2214	2080	1956	1725	1468	1235	1032	805	548	335
Aliskiren	2340	2124	2073	1918	1690	1454	1211	1015	788	564	338
Enalapril	2336	2222	2093	1950	1711	1454	1208	1010	794	539	336

UANALID	TRTVAR	TRTVL	AVISITN	PARAMCD	ROWCAT1	ANLTP1	ANLTP2	STAT	STATVAL	ANLMETH
<STUDYID>_<RA>_XXX1	TRT01P	Combination		PCE	Combination vs. Enalapril	RESPONSE	Experimental	SMALLN	770	Lifetest KM
<STUDYID>_<RA>_XXX1	TRT01P	Combination		PCE	Combination vs. Enalapril	RESPONSE	Experimental	BIGN	2340	Lifetest KM
<STUDYID>_<RA>_XXX1	TRT01P	Combination		PCE	Combination vs. Enalapril	RESPONSE	Experimental	PERCENT	32.9	Lifetest KM
<STUDYID>_<RA>_XXX1	TRT01P	Enalapril		PCE	Combination vs. Enalapril	RESPONSE	Enalapril	SMALLN	808	Lifetest KM
<STUDYID>_<RA>_XXX1	TRT01P	Enalapril		PCE	Combination vs. Enalapril	RESPONSE	Enalapril	BIGN	2336	Lifetest KM
<STUDYID>_<RA>_XXX1	TRT01P	Enalapril		PCE	Combination vs. Enalapril	RESPONSE	Enalapril	PERCENT	34.6	Lifetest KM
<STUDYID>_<RA>_XXX1	TRT01P			PCE	Combination vs. Enalapril	COMPARISON		Hazard	0.93	Lifetest KM
<STUDYID>_<RA>_XXX1	TRT01P			PCE	Combination vs. Enalapril	COMPARISON		95CILOW	0.846	Lifetest KM
<STUDYID>_<RA>_XXX1	TRT01P			PCE	Combination vs. Enalapril	COMPARISON		95CIHIGH	1.03	Lifetest KM
<STUDYID>_<RA>_XXX1	TRT01P			PCE	Combination vs. Enalapril	COMPARISON		1sidedp	0.0862	Lifetest KM
<STUDYID>_<RA>_XXX1	TRT01P			PCE	Combination vs. Enalapril	COMPARISON		2sidedp	0.1724	Lifetest KM
<STUDYID>_<RA>_XXX1	TRT01P			PCE	Combination vs. Enalapril	COMPARISON		adj1sidep	0.3448	Lifetest KM
<STUDYID>_<RA>_XXX1	TRT01P	Combination		PCE	Non Diabetes: Combination vs Enalapril(1)	RESPONSE	Experimental	SMALLN		Lifetest KM
<STUDYID>_<RA>_XXX1	TRT01P	Combination		PCE	Non Diabetes: Combination vs Enalapril(1)	RESPONSE	Experimental	BIGN		Lifetest KM
<STUDYID>_<RA>_XXX1	TRT01P	Combination		PCE	Non Diabetes: Combination vs Enalapril(1)	RESPONSE	Experimental	PERCENT		Lifetest KM

# Open challenges: communicating uncertainty

## Communicating uncertainty about facts, numbers and science

Anne Marthe van der Bles, Sander van der Linden, Alexandra L. J. Freeman, James Mitchell, Ana B. Galvao, Lisa Zaval and David J. Spiegelhalter

Published: 08 May 2019 | <https://doi.org/10.1098/rsos.181870>

Editorial

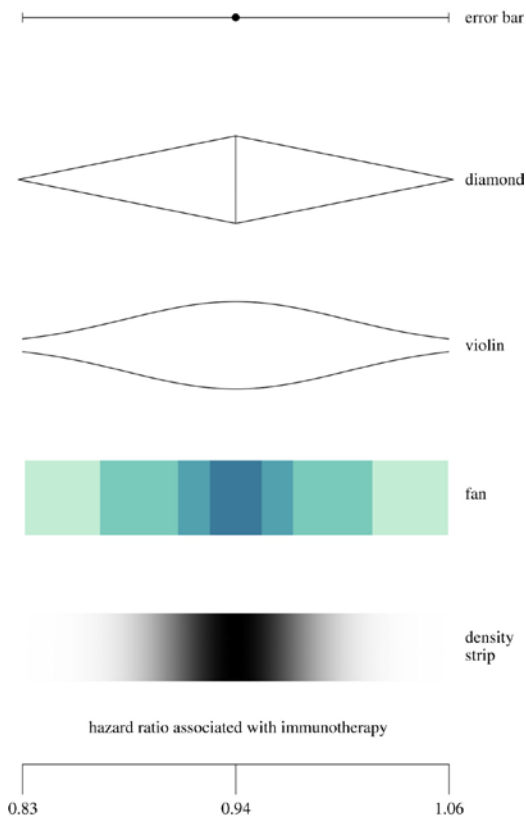
## Moving to a World Beyond “ $p < 0.05$ ”

Ronald L. Wasserstein, Allen L. Schirm & Nicole A. Lazar

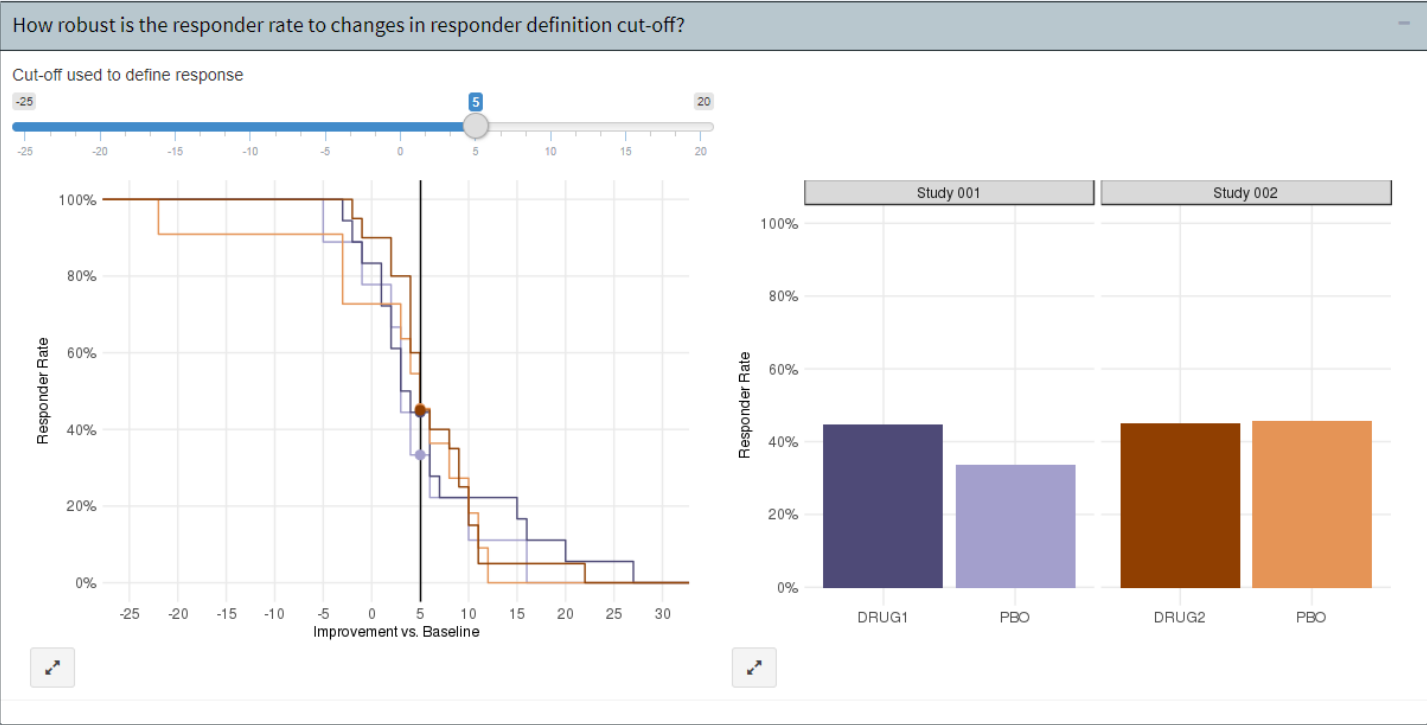
Pages 1-19 | Published online: 20 Mar 2019

Download citation | <https://doi.org/10.1080/00031305.2019.1583913>

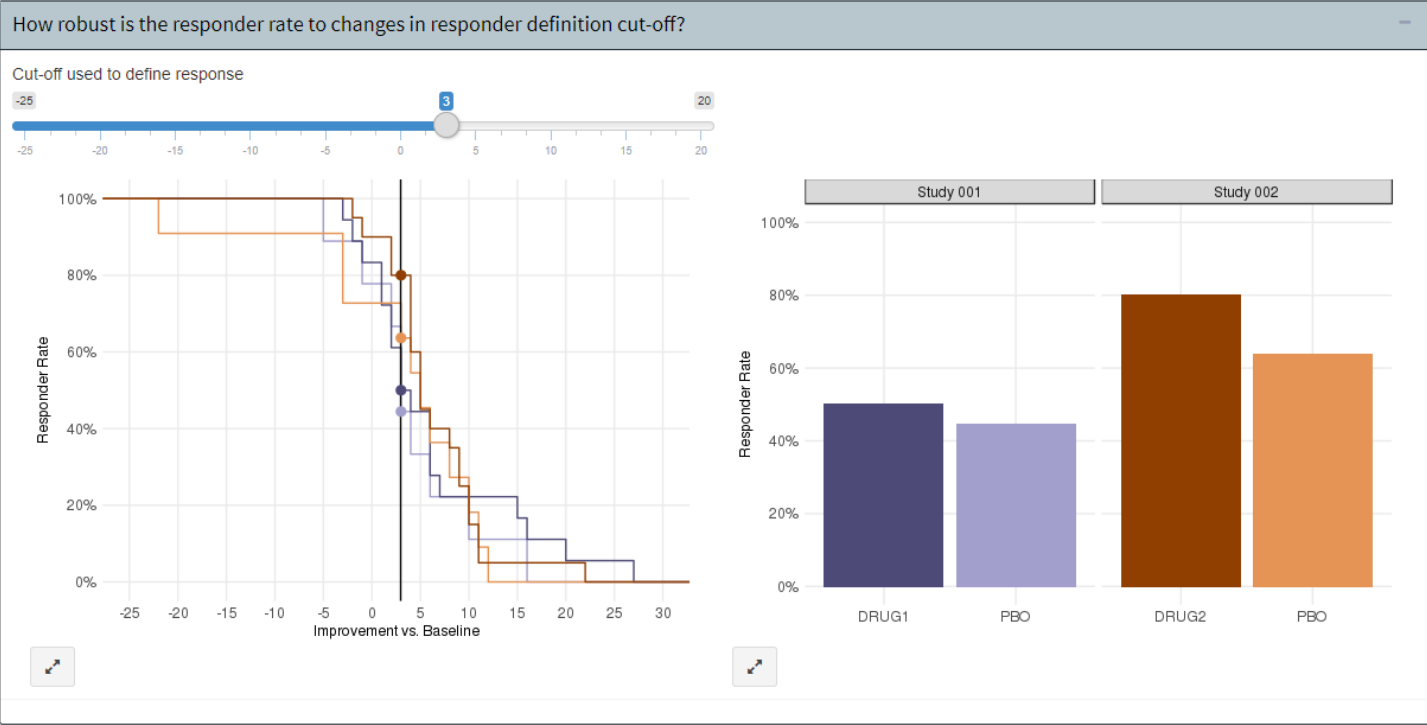
Check for updates



# Analyse questions not data

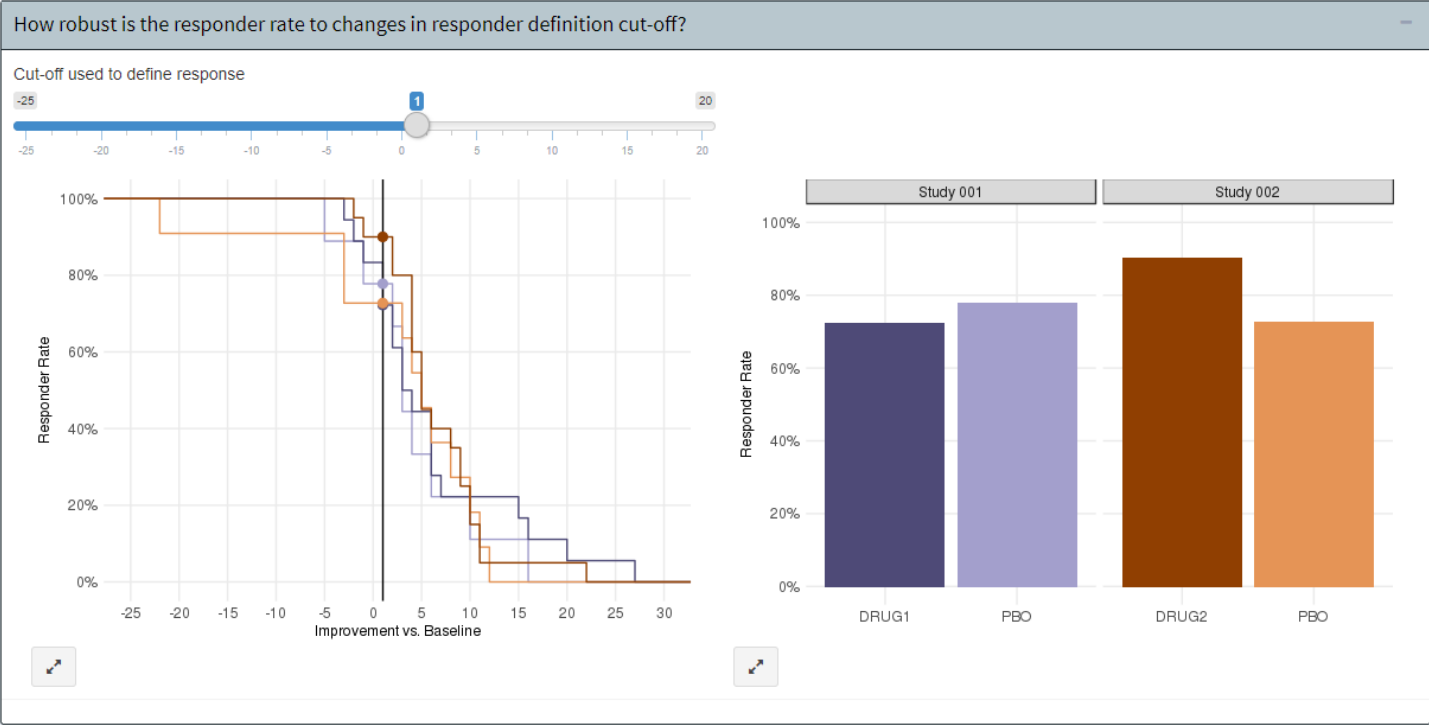


# Analyse questions not data





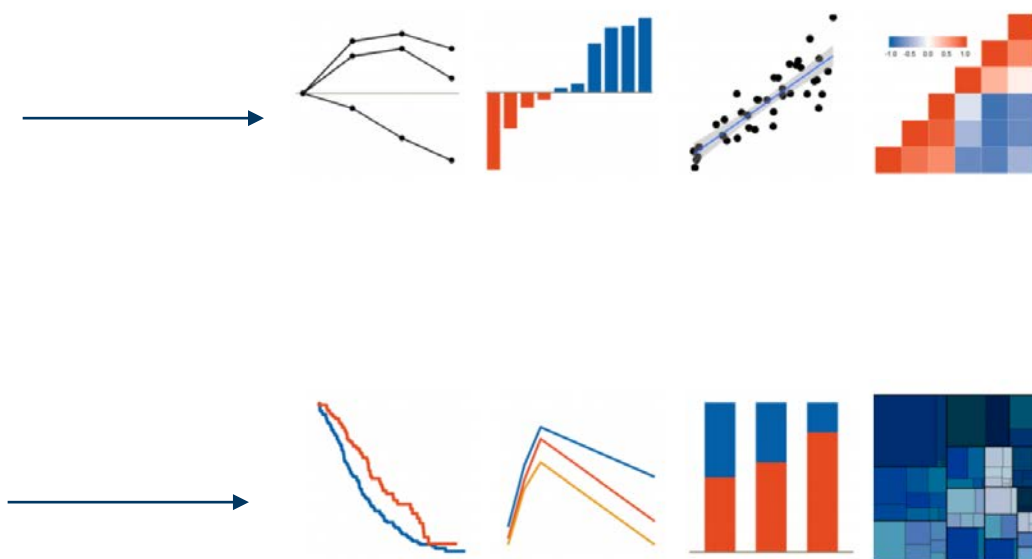
# Analyse questions not data



# Elements of a STRATOS VP initiative

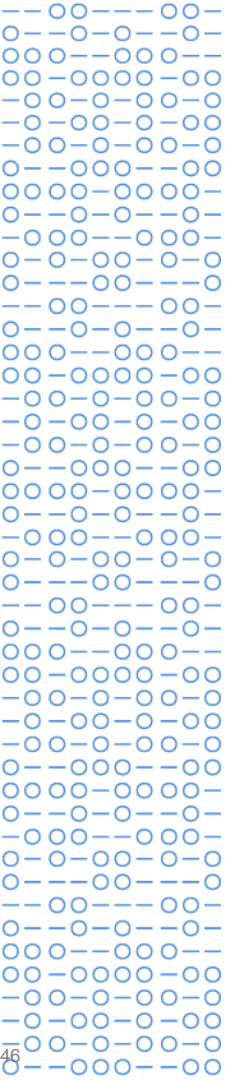
## Topic groups

1	Missing data
2	Selection of variables and functional forms in multivariable analysis
3	Initial data analysis
4	Measurement error and misclassification
5	Study design
6	Evaluating diagnostic tests and prediction models
7	Causal inference
8	Survival analysis
9	High-dimensional data



# Effective data visualisation is effective visual communication

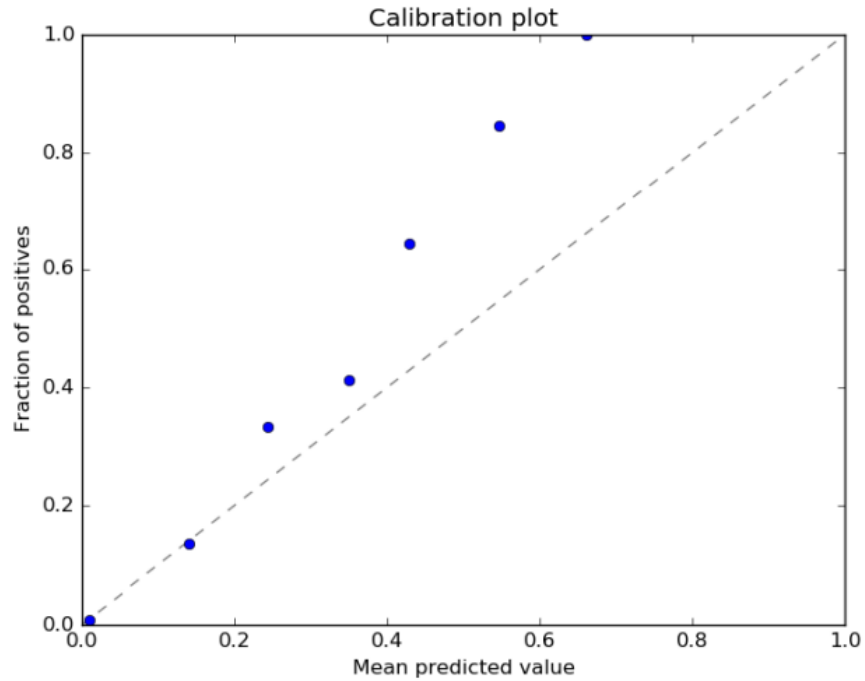
- Effective graphs...
  - are visually appealing, intuitive, legible
  - use the correct graph type and axis scales
  - use proximity & alignment to facilitate comparison
  - use labels and annotations to add clarity to the message
- Most importantly, effective use of visualisations
  - Enables clear and impactful communication
  - Elevates influence with stakeholders
  - Facilitates informed decision making



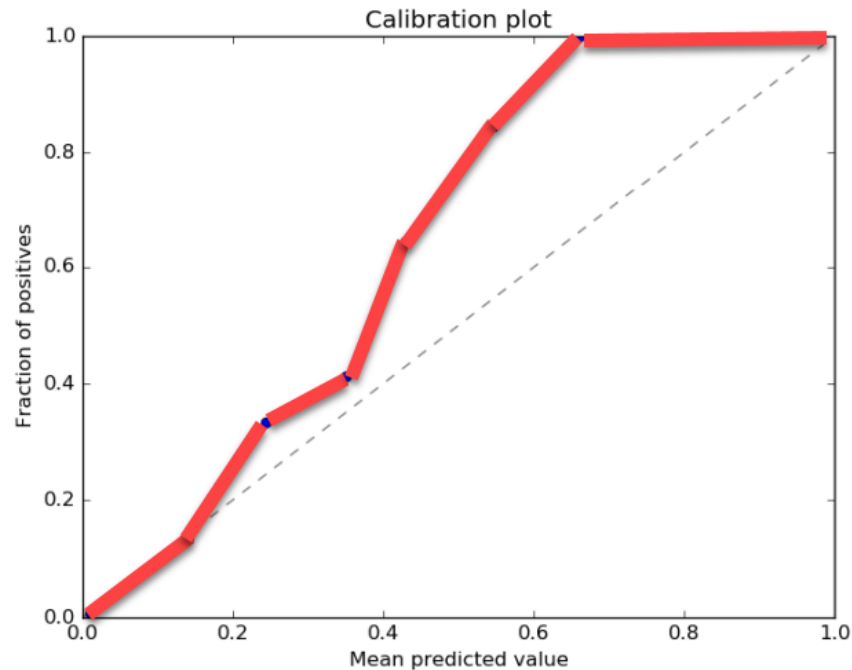
Thank you

# Acknowledgements

- Alison Margolskee
- Marc Vandemeulebroecke
- David Carr
- Andrew Wright
- Jürgen Löffler
- Linda Kanitra
- Baldur Magnusson
- Matthew Brierley
- David Ohlssen
- Shaun Butcher
- Julie Jones
- Walter Hufford
- Ruquan You
- Ivan-Toma Vranesic
- Ian Rees
- Nicolas Guerro
- Keo Chanthavinout
- Frank Bretz



(a) Calibration curve for inpatient mortality predicted at 24 hours into hospitalization for hospital A



(a) Calibration curve for inpatient mortality predicted at 24 hours into hospitalization for hospital A