Visualising harms in publications of randomised controlled trials: consensus and recommendations

Rachel Phillips,[1,2] Suzie Cro,[1] Graham Wheeler,[1] Simon Bond,[3] Tim P Morris,[4] Siobhan Creanor,[5] Catherine Hewitt,[6] Sharon Love,[4] Andre Lopes,[7] Iryna Schlackow,[8] Carrol Gamble,[9] Graeme MacLennan,[10] Chris Habron,[11] Anthony C Gordon,[12] Nikhil Vergis,[13] Tianjing Li,[14] Riaz Qureshi,[14] Colin C Everett,[15] Jane Holmes,[16] Amanda Kirkham,[17] Clare Peckitt,[18] Sarah Pirrie,[17] Norin Ahmed,[19] Laura Collett,[20] Victoria Cornelius[1]

For numbered affiliations see end of the article

Correspondence to: R Phillips
r.phillips@imperial.ac.uk
(or @RachPips on Twitter;
ORCID 0000-0002-3634-7845)

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2022;377:e068983
http://dx.doi.org/10.1136/bmj-2021-068983

Accepted: 05 April 2022

ABSTRACT

OBJECTIVE

To improve communication of harm in publications of randomised controlled trials via the development of recommendations for visually presenting harm outcomes.

DESIGN

Consensus study.

SETTING

15 clinical trials units registered with the UK Clinical Research Collaboration, an academic population health department, Roche Products, and *The BMJ*.

PARTICIPANTS

Experts in clinical trials: 20 academic statisticians, one industry statistician, one academic health economist, one data graphics designer, and two clinicians.

MAIN OUTCOME MEASURES

A methodological review of statistical methods identified visualisations along with those recommended by consensus group members. Consensus on visual recommendations was achieved (at least 60% of the available votes) over a series of three meetings with participants. The participants reviewed and critically appraised candidate visualisations against an agreed framework and voted on whether to endorse each visualisation. Scores marginally below this threshold (50-60%) were revisited for further discussions and votes retaken until consensus was reached.

RESULTS

28 visualisations were considered, of which 10 are recommended for researchers to consider in publications of main research findings. The choice of visualisations to present will depend on outcome type (eg, binary, count, time-to-event, or continuous), and the scenario (eg, summarising multiple emerging events or one event of interest). A decision tree is presented to assist trialists in deciding which visualisations to use. Examples are provided of each endorsed visualisation, along with an example interpretation, potential limitations, and signposting to code for implementation across a range of standard statistical software. Clinician feedback was incorporated into the explanatory information provided in the recommendations to aid understanding and interpretation.

CONCLUSIONS

Visualisations provide a powerful tool to communicate harms in clinical trials, offering an alternative perspective to the traditional frequency tables. Increasing the use of visualisations for harm outcomes in clinical trial manuscripts and reports will provide clearer presentation of information and enable more informative interpretations. The limitations of each visualisation are discussed and examples of where their use would be inappropriate are given. Although the decision tree aids the choice of visualisation, the statistician and clinical trial team must ultimately decide the most appropriate visualisations for their data and objectives. Trialists should continue to examine crude numbers alongside visualisations to fully understand harm profiles.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Harm outcomes data are complex, but visualisations can provide a clear summary of the harm profile and help identify potential adverse drug reactions

Reporting data for harm outcomes in clinical trial manuscripts can be suboptimal

Researchers have requested guidance on appropriate visualisations for harm outcomes and case studies detailing examples of use

WHAT THIS STUDY ADDS

To aid researchers in their choice of visualisations, this study undertook a consensus and endorsed visualisations, presented alongside a decision tree, to communicate harms in the randomised controlled trial setting that can be used as alternatives to the widely used contingency tables

The choice of visualisation will depend on outcome type (eg, binary, time-to-event), scenario (eg, summarising multiple emerging events), trial design (trials with >2 treatment groups require more care), and purpose of the plot (eg, to communicate information about the entire harm profile)

Increasing the use of visualisations will provide clearer presentation of information on harm outcomes and thus enable informative interpretation, especially for assessing the harm profile

Introduction

Well designed graphics are an effective way of communicating messages to a range of audiences and help to identify patterns in data that might otherwise be missed.[1] In 1983, Tufte stated, "of all methods for analysing and communicating statistical information, well-designed graphics are usually the simplest and at the same time the most powerful."[2] In clinical trials, when analysing emerging harm outcomes (ie, non-prespecified events that are reported during the trial and might be unexpected) for which a lot of complex data are collected, visualisations can help to summarise harm profiles (ie, the summary or burden of the cumulative effect of all harm outcomes) and identify potential adverse (drug) reactions. Adverse drug reactions are defined as harm outcomes where a causal relationship between the intervention and event is "at least a reasonable possibility."[3 4] Trials can also prespecify events as harm outcomes of interest to follow up. Prespecified events are individual events that are listed in advance as harm outcomes of interest. These events might be known or suspected to be associated with the intervention or be followed-up for reasons of interest, and visualisations can be beneficial here too. Trial reporting guidelines encourage the use of visualisations for exploring harm outcomes, including the CONSORT (consolidated standards of reporting trials) extension to harms, the 2016 recommendations to improve adverse event reporting from industry representatives and journal editors, a pharmaceutical industry standard from the Safety Planning, Evaluation, and Reporting Team, and guidance from regulators on statistical principles in clinical trials (known as ICH E9).[5-8] The term adverse event is used interchangeably in the literature to refer to harm outcomes but is technically defined as "any untoward medical occurrence that may present during treatment with a pharmaceutical product but which does not necessarily have a causal relationship with this treatment."[3] Potential visualisations for harm outcomes are in abundance but their use in journal articles is limited.[5 6 9 10] A systematic review from 2018 found that only 12% of journal articles made use of visual summaries for adverse event data; a finding supported by a 2019 survey of the UK Clinical Research Collaboration clinical trial unit statisticians.[11 12] However, a 2016 survey of pharmaceutical industry statisticians suggested that in-house practice in this sector might differ.[13] Evidence suggests that journal articles tend to summarise harm outcomes from randomised controlled trials in simple tables of frequencies and percentages, despite the advantages that visualisations offer.[14]

Advances in computer software have improved trialists' capability to produce visualisations; however, little guidance exists on what and how to visually display complex harm data in journal articles. This has resulted in independent calls from the statistical community for direction on "how to decide which of many possible graphics to draw."[12 15] Therefore, with a range of visualisation options available and the increasing ease with which they can be implemented, we sought a consensus to support researchers in their choice of visualisations for randomised controlled trial publications. In collaboration with the UK Clinical Research Collaboration clinical trial unit Statistics Operations Group, we provide recommendations on which visualisations researchers should consider using in the publication of their main research findings.

Methods

We held a series of consensus meetings with 20 statisticians from 15 UK Clinical Research Collaboration registered clinical trial units, one health economist based at an academic population health department, one industry statistician, and one data graphics designer who is part of the multimedia team at *The BMJ*. All these participants are experienced clinical trialists or have an interest in the visual representation of data, or fit into both categories. Against an agreed framework, the group reviewed and critically evaluated 28 plots proposed for visualising data for harm outcomes and refined these plots as necessary, predominantly focusing on clinical trials of an investigational medicinal product. Examples of each of the candidate plots was produced by use of data from one of four completed parallel arm randomised controlled trials and a synthetic dataset (see supplement 1 for further details). The group sought consensus on the plots to endorse and then developed recommendations. To support researchers analysing and interpreting harm outcomes, we present a decision tree to aid their choice of visualisations. We focused on static plots that allow a comparison between treatment groups, in line with the aims of randomised controlled trials that make such inferences. Supplement 1 provides details of the methods used for identification of the considered plots, the consensus process, and how the recommendations were developed. In this paper, we describe each of the endorsed plots, give an example interpretation, and provide our recommendation.

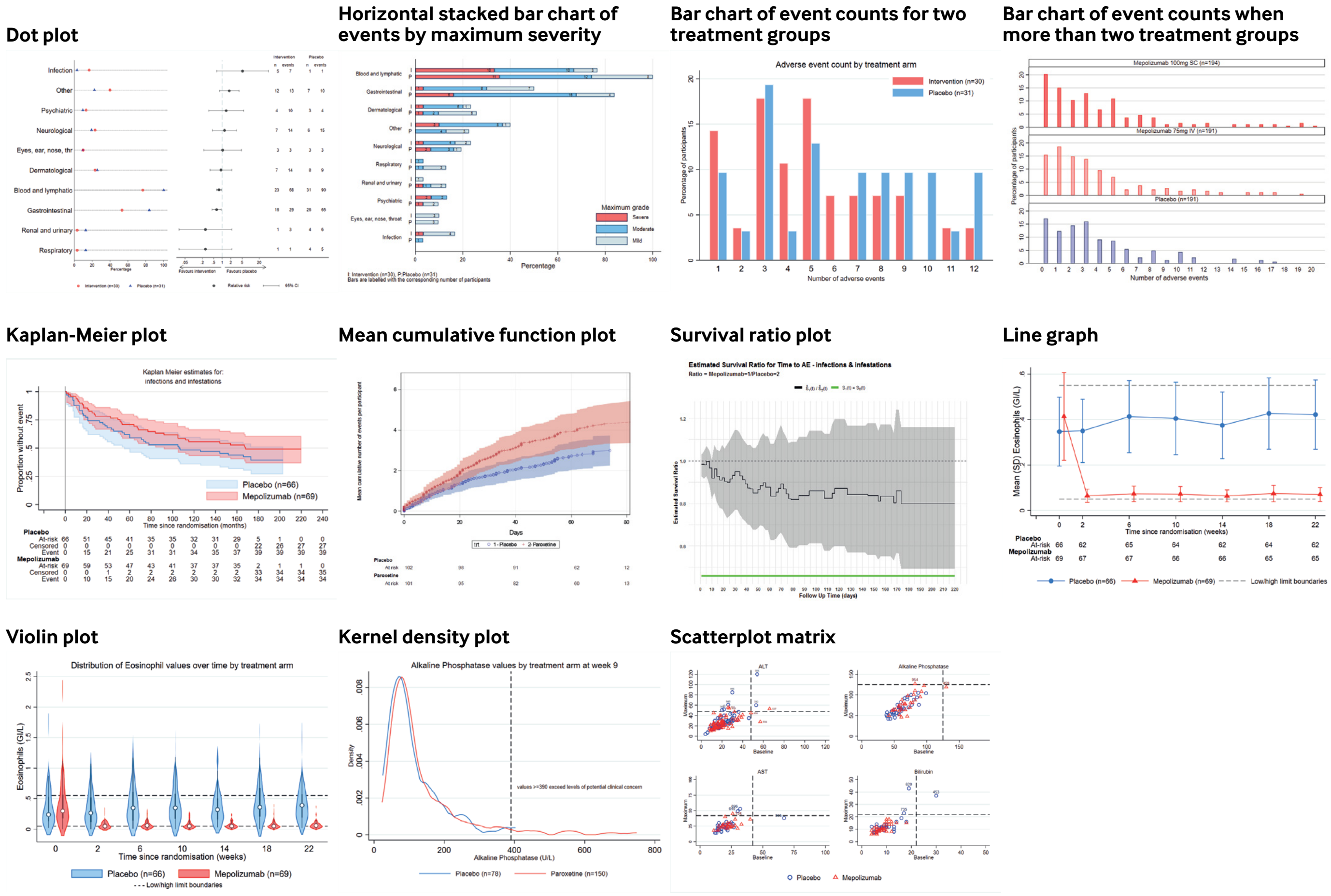
Patient and public involvement

This work forms part of a wider research project that was developed with input from a range of patient representatives. No patient representatives were directly involved in this work, but representatives with experience as clinical trial participants and patient and public involvement advisors reviewed the original proposal and patient and public involvement strategy. We did not speak to patients directly for this research because our focus was to identify the best plots to present in scientific journals with a predominantly scientific readership. The next step is to ask patients for feedback.

Results

We provide the endorsed visualisations that researchers should consider using in the publication of their main research findings according to outcome type and number of events (either single outcomes or multiple outcomes simultaneously; fig 1 and for full size images see supplement 2 figs A1-10).

Outcome type includes binary harm outcomes, which includes events such as occurrence of a headache or experiencing nausea, count outcomes (ie, the number of occurrences of an event, such as number of headaches experienced over follow-up), time-to-event outcomes (eg, time from treatment to headache), and continuous outcomes (eg, individual results from a blood count). We present endorsed visualisations, according to whether the entire harm profile is assessed or a direct message conveyed about a particular event or events of interest, alongside recommendations for use (table 1). To help trialists decide on which visualisation to use, a decision tree (fig 2) and a summary table of required outcome characteristics



Dot plot

Horizontal stacked bar chart of events by maximum severity

Bar chart of event counts for two treatment groups

Bar chart of event counts when more than two treatment groups

Kaplan-Meier plot

Mean cumulative function plot

Survival ratio plot

Line graph

Violin plot

Kernel density plot

Scatterplot matrix

Fig 1 | Endorsed visualisations

(table 2) are provided. Researchers should use these tools when specifying their statistical analysis plan to decide which visualisation they will use, for both prespecified and emerging harm outcomes. Eighteen visualisations were considered but not endorsed (see supplement 3 figs A11-28 for descriptions and potential adaptations discussed).

Recommendations for multiple binary outcomes

Dot plot

Plot description

The dot plot summarises both the absolute and the relative risk for multiple events (fig 1, supplement 2 fig A1). The left panel displays the percentage of participants who had an event (labelled on the vertical axis) in each treatment group. The central panel displays a measure of comparison—in our example, the relative risk of observing each event in the treatment group compared with the control group is shown, along with corresponding 95% confidence intervals on the log10 scale and a line to show the value of no difference (for relative risks, this is 1). Events on the vertical axis are ordered with the highest risk at the top and decreasing in relative risk at the bottom of the graph. The 95% confidence interval shows the uncertainty around the comparative estimate, and its proximity relative to the value of no difference indicates the strength of evidence against the null hypothesis of no difference in event risk between treatment and control groups. The right panel displays a data table containing the number of participants with at least one event and the number of events by treatment group.

Implementation and interpretation

In our example (fig 1, supplement 2 fig A1), the overall impression is that point estimates for the relative summary statistic are evenly distributed on either side of the vertical line but with great differences in levels of precision, shown by the length of the confidence interval, due to the marked differences in the frequencies of the outcome. The largest relative risk communicates increased risk of infection in the intervention group, but the absolute risk and frequencies in the data table show small numbers of participants who had this event. The data show a reduced risk of respiratory events and renal and urinary events in the intervention group; again, the absolute risks and the raw numbers in the data table show only small numbers who had these events. Of note are the estimates for blood and lymphatic disorders and gastrointestinal events, where the relative risks indicate a reduced risk in the intervention group with confidence intervals that do not cross 1. Although these estimates look small compared with

Table 1 | Endorsed plots and recommendations for use

Outcome type	Plot	Recommendation
Visualisations for summarising entire harm profile (viewing differing multiple adverse events)		
Binary	Dot plot	Use to present a comprehensive summary of the occurrence of multiple binary events
Binary	Stacked bar chart	Use to present information on the occurrence and severity of multiple binary events
Count	Bar chart	Use to present information on event counts
Continuous	Scatterplot matrix	Use in an exploratory setting to help identify any outliers or patterns of interest across multiple continuous outcomes
Time to event	To be developed	No plot endorsed
Visualisations to summarise an event of interest* (viewing a single adverse event)		
Time to event	Kaplan-Meier plot with extended at-risk tables	Use to present information for specific events of interest and to detect either a large between treatment group difference or potential disproportionality over time
Time to event	Survival ratio plot	Use as a signal detection tool to spot departures from unity to help detect potential signals for adverse drug reactions, and alongside the Kaplan-Meier plot to incorporate a direct estimate of between group difference for time-to-event outcomes
Time to event	Mean cumulative function plot	Use to display time-to-event information for recurrent events. Provides a visual summary of the time to expect a certain number of an event to be experienced per participant by treatment group
Continuous	Line graph	Use to describe continuous harm outcomes of interest over time, using an appropriate summary statistic including an indication of variability
Continuous	Violin plot	Use as an alternative plot to the line graph to present a description of continuous harm outcomes of interest over time if, for example, the outcome of interest is far from a normal distribution and/or there is interest in exploring the distribution
Continuous	Kernel density plot	Use to explore and compare an outcome of interest at a specific time point or to investigate how an outcome of interest changes from baseline to either a specific point in time or maximum change over the entire trial period

*Where an event can be a single adverse event (eg, headache) or a single category of events that have been grouped together (eg, neurological body system) or an aggregated summary (eg, number of serious adverse events).

the other relative risks, the left side of the plot clearly shows a noticeable difference in absolute numbers, and the data table shows the large numbers of patients who had these events. Therefore, this finding suggests a potential beneficial effect of the intervention on these harm outcomes that might warrant closer inspection.

Recommendation

The consensus group unanimously endorsed the dot plot for presenting data for multiple binary outcomes. The dot plot provides a comprehensive presentation of the data that incorporates the traditional table of events. The dot plot was the only visualisation to receive 100% endorsement (see supplement 6 for the endorsement consensus for the other recommended plots).

Potential amendments

The relative risk, risk difference, odds ratio, or incident rate ratios (adjusted or unadjusted as desired) can be plotted as the measure of comparison in the central panel of this plot. Some researchers might also prefer

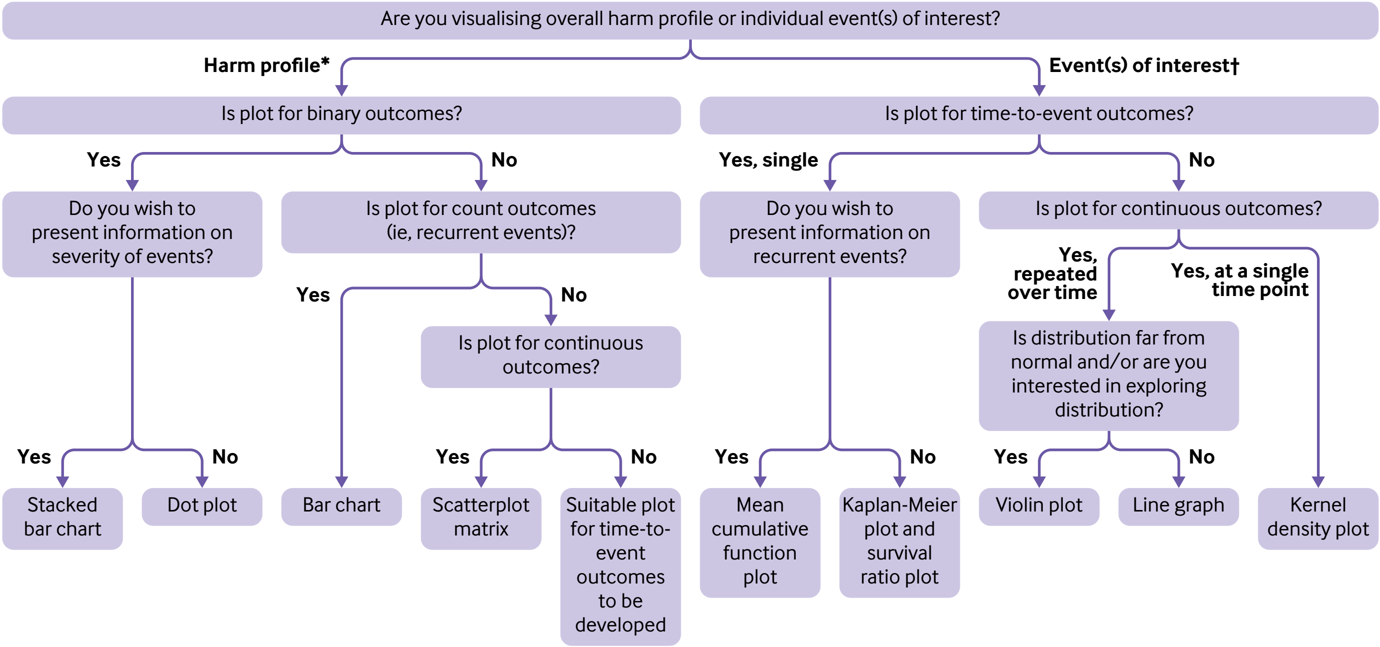

Are you visualising overall harm profile or individual event(s) of interest?
Harm profile*
Event(s) of interest†
Is plot for binary outcomes?
Yes
No
Do you wish to present information on severity of events?
Is plot for count outcomes (ie, recurrent events)?
Yes
No
Is plot for continuous outcomes?
Yes
No
Stacked bar chart
Dot plot
Bar chart
Scatterplot matrix
Suitable plot for time-to-event outcomes to be developed
Is plot for time-to-event outcomes?
Yes, single
No
Do you wish to present information on recurrent events?
Is plot for continuous outcomes?
Yes, repeated over time
Yes, at a single time point
Is distribution far from normal and/or are you interested in exploring distribution?
Yes
No
Mean cumulative function plot
Kaplan-Meier plot and survival ratio plot
Violin plot
Line graph
Kernel density plot

Fig 2 | Decision tree to support selection of plots to visualise data for harm outcomes. *Summary of all harm outcomes collected. Individual events include individual emerging events (including adverse events and laboratory or vital sign data indicative of harm) and prespecified events of interest. †Can include a single adverse event (eg, headache), a single category of events that have been grouped together (eg, neurological body system), or an aggregated summary (eg, number of serious adverse events)

Table 2 | Summary of characteristics to guide researchers in their choice of plot to visualise data for harm outcomes

Characteristics of outcome to be displayed	Plot
Binary	
Multiple outcomes	Dot plot
Multiple outcomes with severity ratings	Stacked bar chart
Count (recurrent) outcome	Bar chart
Continuous	
Multiple outcomes	Scatterplot matrix
Single outcome, repeated over time	Line graph
Single outcome, repeated over time with non-normal distribution or interest in exploring the distribution	Violin plot
Single outcome, at single time point	Kernel density plot
Time-to-event	
Multiple outcomes	No suitable plot
Single outcome	Kaplan-Meier plot and survival ratio plot
Single, recurrent outcome	Mean cumulative function plot

to present the data table in the central panel so that it appears alongside the absolute summary. This plot can be presented in grayscale without any loss of meaning. A small number of additional arms can be added for multiarm studies through incorporation of multiple non-overlapping estimates on the same plot (eg, by use of jittering); however, an increase in the number of active treatment groups can lead to incomprehensive distinction between arms.

Limitations

Confidence intervals around the relative differences are useful to identify potential signals (that is, information that raises the possibility of a causal relationship between the intervention and event) of harm for further investigation. However, confidence intervals should not be used as a proxy for hypothesis testing, which will increase the chance of finding spurious significant differences resulting from multiple hypothesis tests.[16] Clinician feedback indicated that trialists should consider varying the horizontal axis range for the absolute summary and scale for the relative summary to ensure clarity without exaggerating effects—for example, presentation of the entire 0-100 scale for the absolute summary might not be appropriate for rare events. When presenting the odds ratios or risk ratios, if no events were reported in one of the treatment groups, a common, simple correction is to add half an event to each group (numerator and denominator). This continuity correction is commonly used but has been shown to be inferior when undertaking meta-analyses for rare events; therefore, alternative corrections might warrant consideration.[17 18] Although this plot gives a comprehensive overview, some potentially important pieces of information are not included, such as the relative severity of different harm outcomes, and even though recurrent events can be presented using the incident rate ratio, this information cannot be easily displayed on the left panel. In scenarios where information on severity is important, the stacked bar chart can be used, and for recurrent events, the mean cumulative function plot can be used (see later).

Software

The dot plot can be produced in Stata by use of the aedot or aedots command, in R with the code available in supplement 4, and in SAS with the code available from the CTSpedia Wiki page (https://www.ctspedia.org/do/view/CTSpedia/ClinAEGraph000). The SAS example does not include code to incorporate the data table.[19]

Stacked bar chart

Plot description

The horizontal stacked bar chart presents the percentage of participants with an event by treatment group and by maximum severity—that is, if a participant had the same event twice, once classified as mild and once as moderate, this participant would be counted once as having a moderate event (fig 1, supplement 2 fig A2). The bars are labelled with the corresponding number of participants. Bars are split by colour gradient to indicate different severity grades, and the total bar height shows the proportion of participants who have had that event at least once. The most severe grade is displayed closest to the vertical axis to allow ease of informal comparison across treatment groups for the most harmful or burdensome events.

Implementation and interpretation

In our example (fig 1, supplement 2 fig A2), the most frequent events reported were at least one event of the blood and lymphatic system or gastrointestinal disorders. Although more blood and lymphatic events were noted in the placebo group, the stacked bar chart shows that the proportion in the most severe categories (severe plus moderate) were similar across treatment groups, and the difference in numbers between treatment groups was because of the difference in participants who had a mild event. For gastrointestinal disorders, the stacked bar chart showed that fewer events were recorded for the intervention group across each of the severity grades compared with those in the placebo group. The plot also displays events classified as other that were dominated by severe and moderate events in the intervention group compared with the placebo group, which could warrant closer inspection of the type of events. The stacked bar chart highlights the most frequent events because of the increased physical space that these events occupy. This display contrasts with the dot plot, in which the most frequent events take up the least space in the central panel because of the increased precision and hence narrower confidence intervals around the treatment effect estimate.

Recommendations

The stacked bar chart is easy to understand and is useful when it is important to present information on severity of multiple events. This display can be used to informally compare severe or severe plus moderate events or the overall number of events between groups. Treatment groups are recommended to be displayed directly adjacent to each other for each event and horizontally aligned to allow labelling that is easy to read.

Potential amendments

This plot can be adapted to multiarm studies, and graduation from black to white is possible without loss of meaning to avoid use of other colours. The single event setting can make use of this graph by replacing events on the vertical axis with representation of time—for example, visits or treatment cycle, an example of which can be found in Thanarajasingam et al.[20]

Limitations

Direct comparisons of stacked bars within severity ratings between treatment groups are not possible beyond the segment closest to the vertical axis; however, cumulative comparisons such as severe plus moderate are possible and are perhaps more meaningful. Stacked bar charts promote presentation of information on participants with at least one event at maximum severity rather than number of events, and additional information on repeated events should also be presented. In addition, the effect sizes for differences between groups are not explicitly displayed.

Software

Stacked bar charts are easily implemented as standard plots across the variety of statistical packages (graph hbar, Stata; barplot or the ggplot2 package with geom_bar, R; proc gchart, SAS).

Recommendations for single binary outcomes

Bar chart

Plot description

A bar chart presents information on the number or count of adverse events reported per participant (fig 1, supplement 2 figs A3a and b). Each bar represents the percentage of participants by number of events experienced for each treatment group.

Implementation and interpretation

Figure 1 (supplement 2 fig A3a) displays the distribution of multiple events, with higher numbers of multiple events recorded more often for participants in the placebo group than participants in the intervention group. In the alternative figure (fig 1 supplement 2 fig A3b) the distributions indicate that participants in either of the intervention groups had multiple events more often than those in the placebo group.

Recommendations

The bar chart is recommended to present information on the number of events experienced. This plot is simple and can be useful to illustrate differences in counts of binary events between treatment groups and is potentially useful to highlight differences in the burden of harm experienced by participants. A bar chart can depict an overall summary of events, such as the total number of serious adverse events, a limited number of events of interest, or a single event of interest. This plot can also be used in an exploratory setting to show the distribution of repeated events.[21 22] Vertical bars with treatment groups presented alongside each other are the recommended format (fig 1 supplement 2 fig 3a) when comparing two treatment groups. For more than two treatment groups, the recommended alternative is to use separate plots stacked above each other for each group (fig 1 supplement 2 fig 3b).

Potential amendments

This plot can be easily adapted to multiarm studies and can be produced in grayscale if necessary. Additionally, bars could be labelled with number of participants to ensure accurate communication.

Limitations

Although this plot is helpful for summarising and comparing the overall burden of different treatments, it does not make a distinction between the types of events. Therefore, trialists should still explore and report the individual event data, giving careful consideration as to whether such a plot for overall events could be misleading. In addition, although bar charts could potentially reveal patterns in the data, clinician feedback indicated that subtle differences would be less obvious, and careful consideration of when to use this plot and the accompanying message it supports is needed.

Software

Bar charts are easily implemented as standard plots across the variety of statistical packages (graph bar, Stata; barplot or the ggplot2 package with geom_bar, R; proc gchart, SAS).

Recommendations for single time-to-event outcomes

Kaplan-Meier plot

Plot description

The Kaplan-Meier plot for single time-to-event outcomes shows the cumulative proportion of participants remaining event-free over time by treatment group (fig 1, supplement 2 fig A4). The 95% confidence interval bands indicate the precision of the within group estimates of being event-free. The table below the plot shows the number of participants who remain at risk for the specific event of interest, the cumulative number who have been censored, and the cumulative number who had the event of interest at each discrete time point.

Implementation and interpretation

In our example the extended risk table (fig 1) indicates that by the end of follow-up, little difference was noted between treatment groups in the number of participants who had an infection or infestation. However, the event curves show that 50% of the placebo group had this event within about 100 days of randomisation, whereas it took until 160 days after randomisation for 50% of the mepolizumab group to experience the event.

Recommendations

We recommend the Kaplan-Meier plot with within group confidence intervals and extended risk table for specific events of interest to detect either a large between treatment group difference or a potential disproportionality over time, especially as adverse drug reactions are often time dependent.

Potential amendments

For rare events, trialists might want to reverse the vertical axis to display the cumulative proportion with the event to aid interpretation. This plot can be created in grayscale, with different line styles used to differentiate between groups. Extensions to multiple events or multiarm studies are potentially feasible but can become incomprehensible when displaying multiple overlying confidence bands. Therefore, trialists should consider only plotting the survival estimates with extended risk tables, or present separate plots for comparison of each intervention group, with a common comparator or separate plots for different events.

Limitations

Kaplan-Meier plots depict only time-to-first event, failing to consider recurrent events. For clarity in presentation, these graphs are also typically limited to one type of event at a time. To present information on recurrent events over time, a plot of the mean cumulative function (see later) is recommended. Some generic limitations of using time-to-event plots in this setting are discussed later.

Software

Kaplan-Meier plots are easily implemented as standard plots across a variety of statistical packages. To incorporate the extended risk tables, trialists can use the R package KMunicate and a program for implementation in Stata (https://github.com/sarwarislam/kmunicate_stata).[23]

Mean cumulative function plot

Plot description

For recurrent events or a summary of the total burden of events, the mean cumulative function plot is recommended. This plot is a non-parametric estimate of the mean cumulative number of events per participant (displayed on the vertical axis) as a function of time (horizontal axis) by treatment group (fig 1, supplement 2 fig A5). The 95% confidence interval bands show the precision of the within group estimate. The risk table includes information on the number of participants who remain at risk of an event at discrete time points.

Implementation and interpretation

Over the first week after randomisation, the mean number of events per participant is similar across treatment groups, but by day 20 a divergence becomes apparent (fig 1). In the paroxetine group, a mean of two events per participant was observed by day 20, but in the placebo group at that time a mean of approximately 1.5 events per participant was observed. The plot of the mean cumulative function shows the participant burden of recurrent events, highlighting in this example that over follow-up, participants in the paroxetine group had on average a greater number of events than participants in the placebo group, suggesting that some events are associated with the intervention.

Recommendations

Unlike the Kaplan-Meier plot, this plot can display information on recurrent events, providing a visual summary of the expected time until a certain number of an event will be recorded per participant by group. This visualisation can show the burden of any event as in the example that we present, or the recurrence of events of special interest. As highlighted in the clinician feedback, these plots are potentially useful when investigating long term treatments for chronic conditions and can provide valuable insight into periods when the treatment might be considered safe or well tolerated. When used to present data for any event, this plot serves as an alternative to the bar chart of counts that incorporates time. This graph also usefully summarises overall burden in place of, or in addition to, summaries of time to discontinuation that are often reported as a proxy for harm.

Potential amendments

As with the Kaplan-Meier plot, this plot can be created in grayscale without loss of meaning. Extension to multiarm studies or multiple events is potentially feasible, but displaying multiple overlying confidence bands could make the plot incomprehensible. Similar to the recommendation for the Kaplan-Meier plot, trialists should therefore consider plotting only the mean cumulative function (without confidence bands) and risk table, or present separate plots for comparison of each intervention group with a common comparator, or separate plots for different events.

Limitations

For clarity in presentation, mean cumulative function plots are typically limited to one type of event at a time. More generic limitations and cautions of use of time-to-event plots in the harm setting are provided later in this paper.

Software

The mean cumulative function with confidence interval bands can be implemented using the SAS proc reliability procedure and mcfplot command.

Limitations applicable to time-to-event methods

The measure of uncertainty (confidence interval bands) in the Kaplan-Meier plot and the plot of the mean cumulative function are within treatment groups and not between treatment groups, which is the inference of interest in comparative clinical trials. To incorporate an estimate of the between group difference with a measure of uncertainty, the survival ratio plot can be used (see later). Additionally, when time-to-event methods for harm data are used, trialists must remain aware of the limitations around competing risks and consider these when performing the underlying time-to-event analyses. More information on alternative strategies to account for competing risks can be found in Proctor and Schumacher[24] and include use of appropriate estimates (eg, Aalen-Johnson estimator or Fine and Gray method) to plot the cumulative incident function.

Survival ratio plot

Plot description

The survival ratio plot displays the ratio of non-parametric estimates of the survival probabilities (ie, the probabilities for being event-free in the harm setting) between treatment groups over time along with 95% confidence intervals. Unlike the Kaplan-Meier and mean cumulative function plots, this plot allows a direct comparison between treatment groups (fig 1, supplement 2 fig A6). As the plot displays the ratio of survival probabilities over time, departures from unity (point of unity is the survival ratio of 1) indicate potential differences between treatment groups. The green horizontal bar at the bottom of the plot changes to red if the confidence interval excludes unity.[25]

Implementation and interpretation

The survival ratio plot (fig 1) depicts a point estimate indicating a greater risk of infection and infestation in the placebo group compared with the intervention group, with a value between 0.9 and 1.0 until day 40 and dropping below 0.9 thereafter. Compared with the Kaplan-Meier plot, this plot shows the confidence band for the between group comparison (rather than within-group confidence intervals in the Kaplan-Meier plot). The confidence band includes the point of unity across all time periods and therefore would not provide sufficient evidence to raise a signal for this event to undergo further investigation.

Recommendation

The survival ratio plot would be suitable for signal detection analysis for emerging events because it provides a between group comparison that can be used to spot departures from unity and helps to identify the time that such divergences occur, which can help detect potential signals for adverse drug reactions. For events of specific interest when the focus is on accurately estimating survival probabilities over time, this plot is less suitable. This plot can be presented alongside the Kaplan-Meier plot to show both a relative and an absolute measure.

Potential amendments

Our example displays the ratio of survival probabilities estimated using the Kaplan-Meier method; alternatively, the display could show the difference in survival probabilities. As with both the Kaplan-Meier and the mean cumulative function plots, multiple lines can be added to one graph to display estimates for different events or multiple treatment comparisons.

Limitations

As with Kaplan-Meier plots, the survival ratio plot allows for only time-to-first event; therefore, this graph is not suitable for recurrent events. The plot is also limited to one type of event; however, in some situations multiple estimates can be added to the same plot but with the same considerations as plotting multiple lines on the Kaplan-Meier plot. As with other time-to-event plots, competing risks are important to consider when performing the underlying time-to-event analysis, further details of which are discussed above. The confidence interval band of values around the relative differences are useful to detect signals of potential harm for further monitoring, but we are not encouraging hypothesis testing in this setting.[16] Despite survival ratio plots first being proposed in 2006, little evidence exists of their application in the clinical trial literature; use of this plot will need to be accompanied by a detailed explanation until audiences become more familiar with it and its interpretation.[25] This postulation was supported in discussions with clinicians, who initially struggled to interpret this plot but who indicated a strong endorsement after further explanation was provided.

Software

The survival ratio plot can be implemented in R using the survRatio package with the drsurv function to take the time, censoring indicator, and treatment indicator as inputs. This package returns Kaplan-Meier survival estimates and corresponding confidence intervals to create an object of the survival ratio, survival difference, and pointwise (bootstrap) confidence bands. The ggsurv function is then used to create the plot of the survival ratio and confidence interval bands.

Recommendations for single continuous outcomes

Line graph

Plot description

In the line graph plot, markers display mean values and vertical lines indicate the standard deviation (not standard error) of raw values at each discrete time point, connected with a line to the point closest in time for each treatment group (fig 1, supplement 2 fig A7). Horizontal reference lines are included to indicate the upper and lower limits of normal values for the outcome, and a table of numbers of participants at risk at each discrete time point is included.

Implementation and interpretation

An immediate decrease can be seen in the mean eosinophil count after randomisation in the mepolizumab group (fig 1), and this decreased level is maintained across follow-up. The mean values for the placebo group fluctuate around the baseline value and the error bars exceed the upper limit of normal during follow-up.

Recommendations

This plot can be used to describe continuous harm outcomes of interest over time by use of an appropriate summary statistic, together with an indication of variability. This plot can be helpful to identify shifts in distributions between treatment groups and highlight any potential trends; as a result, this display might be better suited to depict clinical outcomes (such as vital signs) rather than blood markers, where clinicians are more often interested in the tails of the distribution (ie, the ends or extremes of the distribution of observed values).

Potential adaptations

The summary statistic displayed in this plot should be chosen to reflect each individual dataset and the purpose of the plot, for example, when interest is in presenting descriptions of the distributions, either means and standard deviations or medians and interquartile ranges can be plotted, and if interest is in drawing inferences of between group comparisons, then estimates from mixed effects models for repeated measures with 95% confidence intervals can be presented. This plot can easily incorporate multiple groups or outcomes and can be modified to exclude the use of colour.

Limitations

Changes in the tails of the distributions are usually of most interest when monitoring blood markers for harm, and such changes might be difficult to see using this plot. This graph is also unsuitable for skewed distributions; alternative plots for such data are presented below. Appropriate colours and line styles should be considered for clarity, particularly when adapting line graphs to multiarm trials.

Software

Line graphs are easily implemented as standard plots across the variety of statistical packages (eg, twoway connected and twoway rspike, Stata; plot and lines or using the ggplot2 package with geom_line and geom_errorbar, R; and proc gplot, SAS).

Violin plot

Plot description

The hollow circle marker on the violin plot indicates the median value, the narrow rectangular boxes indicate the interquartile range, and lines extend from the box to the minimum and maximum points for each group at each time point. These parts are overlaid with kernel density plots (see later), which summarise the distribution of the raw values (fig 1, supplement 2 fig A8). The two horizontal dashed lines indicate the upper and lower limits of normal values.

Implementation and interpretation

At time 0 (randomisation) the distributions were similar across treatment groups, but from week 2 onwards the distribution of values in the mepolizumab group was narrower than in the placebo group (fig 1). The distribution of the values in the placebo group was largely unchanged over time and indicated that a proportion of the participants remained in the upper tail exceeding the upper boundary of normal throughout follow-up. This display indicates a benefit for the mepolizumab group by reducing eosinophil concentrations to within the normal limits.

Recommendation

This plot is an alternative to the line graph to describe continuous data that can be used even if the outcome of interest is not normally distributed. Outlying values are displayed and these can be labelled to highlight participants who persistently record values of concern.

Possible adaptations

In the current format, information is duplicated because the kernel density plot is mirrored. Presenting only one kernel density would improve clarity and produce a more space efficient plot.

Limitations

The violin plot only allows for informal between group comparisons of distributions and does not allow for presentation of formal between group inferences such as the estimates from mixed effects models, which can be presented in a line graph. Adaptations to multiarm trials are not as space efficient as for the line graph. Kernel density estimates for some data might extend to values outside the plausible range—for example, some kernel densities are estimated to be below 0 for eosinophil counts, which is not feasible clinically.

Software

The violin plot can be implemented in Stata by use of vioplot or by use of the ggplot2 package in R with geom_violin or SAS proc sgpanel.

Kernel density plot

Plot description

The kernel density plot displays the distribution of a continuous outcome. Data can be for a single time point or a derived change score—for example, the difference between the baseline value and maximum value while receiving treatment (fig 1, supplement 2 fig A9). Vertical reference lines can be included to indicate the upper and lower limits of normal values for the outcome.

Implementation and interpretation

Although figure 1 shows that values are similarly distributed in the placebo and paroxetine groups when within the normal range (ie, <390 U/L (6.51 µkat/L)), the plot clearly shows a high alkaline phosphate value for some participants in the paroxetine group through the long right tail. This plot highlights the increased alkaline phosphatase concentrations in some participants taking paroxetine as an important event for closer monitoring in future trials or in the postmarketing setting.

Recommendations

The kernel density plot is recommended to explore an outcome of interest at a specific time point or a change score—for example, the change from baseline to a specific point in time or maximum change over the entire trial. This plot can be used to informally compare whole distributions of data between treatment groups and can highlight important differences in these distributions.

Potential adaptations

This plot can easily incorporate multiple groups and can be modified to not require use of colour.

Limitations

The kernel density plot only allows for informal between group comparisons of distributions and the information on repeated measures is lost, only displaying information for one time point.

Software

The kernel density plot can be implemented in Stata by use of twoway kdensity or the ggplot2 package in R with geom_density or SAS densityplot.

Recommendations for multiple continuous outcomes

Scatterplot matrix

Plot description

Multiple scatterplots of continuous outcomes arranged in a matrix, each display the relationship between values at two different time points—for example, baseline values along the horizontal axis and the participant's maximum value over follow-up along the vertical axis (fig 1, supplement 2 fig A10). The dashed lines represent the boundary between normal and abnormal thresholds.

Implementation and interpretation

In our example, where a higher threshold is worse, participants of most concern would be in the top left quadrant (ie, participants' baseline values were normal and are now abnormal) and the participants who have improved would be in the bottom right (ie, participants' baseline values were abnormal and are now normal). If more participants from the intervention group than control group were in the top left quadrant this would be cause for concern. In figure 1, slightly more participants in the placebo group (n=4) had higher alanine transaminase (ALTs) when receiving treatment compared with baseline in contrast with participants in the mepolizumab group (n=2).

Recommendation

The scatterplot matrix is recommended in an exploratory setting to identify any outliers or patterns of interest. We suggest labelling outlying values with a participant identifier, as shown in figure 1, to assess if one or more participants have abnormal measurements across outcomes. This could be useful to monitor participants in ongoing studies and might also help to raise signals for potential adverse drug reactions in final analyses.

Possible adaptations

This plot could be used to explore two continuous measures at any time point over study follow-up. Variations in symbol style and colours should be used to help separate overlapping measurements between groups. Reference lines could be included to indicate both upper and lower limits of normal for each outcome.

Limitations

This plot presents several visual problems. Use of solid colours results in occlusion, making it difficult to distinguish individual points; transparency options could help with this issue.

Software

Scatterplots are easily implemented as standard plots across the variety of statistical packages. For example, use of twoway scatter in Stata to produce the individual plots and the graph combine command or use of the grc1leg command to produce the scatterplot matrix.

Areas for further development

Among the visualisations considered for displaying multiple time-to-event outcomes, the options available were judged to be poor. Although multiple Kaplan-Meier plots could be used to display information on a limited number of prespecified events of interest, a gap remains in how to visualise multiple time-to-event outcomes simultaneously on the same plot. We discussed the development of novel plots in this setting and we will pursue this in future work.

Discussion

Randomised controlled trials provide a valuable source of data to compare harm outcomes between treatment groups and can help to identify potential signals for adverse drug reactions. However, evidence suggests that practices of reporting data for harm outcomes in clinical trial manuscripts are suboptimal. The CONSORT harms extension[5] aimed to improve reporting, and the recommendations from Lineberry et al[6] provided detailed examples to be used alongside the CONSORT harms extension. Both recommendations called for use of visualisations when reporting harm outcomes but did not give guidance on what visualisations would be helpful. Researchers have called for information on appropriate methods of analysing and presenting harm outcomes and for case studies detailing examples of use.[12]

Principal findings

Our aim was to provide consensus recommendations developed over a series of virtual meetings with researchers responsible for producing clinical trial manuscripts, including clinical trial statisticians and researchers from both academia and industry, as well as clinicians. We have provided examples of the endorsed visualisations to communicate risks of harms in the randomised controlled trial setting that can be used as an alternative to the widely used contingency tables. Our purpose was to increase the use of visualisations for harm outcomes in clinical trial manuscripts and reports and promote presentation of clearer and more informative information on harm outcomes to aid interpretation. Each of the endorsed visualisations can be constructed in standard statistical software and we have signposted to accessible code, when available, for implementation, with the aim of supporting adoption and to ensure efficient application of the recommendations. Trialists can implement our recommendations alongside the CONSORT harms extension[5] and the recommendations of Lineberry et al,[6] as well as the more general guidance on the content of statistical analysis plans from Gamble et al.[26]

The choice of visualisation will depend on the outcome type (eg, binary, count, time-to-event, or continuous), the scenario (eg, summarising multiple emerging events or one event of interest), the trial design (trials with >2 treatment groups require more care), and the purpose of the plot (eg, to communicate information about the entire harm profile or to convey a direct message about a particular event of interest). It is for the statistician and clinical trial team to decide the most appropriate visualisation or visualisations for their data and objectives. A combination of plots is likely to be necessary—for example, presenting the traditional Kaplan-Meier plot alongside the survival ratio plot for prespecified harm outcomes to explore the temporal relationship, in addition to the dot plot to summarise the overall harm profile. Researchers can use the decision tree (fig 2) to support their choices, but this tool is not suitable for all circumstances; consideration is still required when deciding the most appropriate visualisations. Different metrics will need to be used depending on what is important to show. For example, for continuous outcomes some of the plots include the standard deviation, which measures the amount of variability of individual data from the sample mean, some include the standard error, which is a measure of precision of the sample mean, and others include the 95% confidence interval, which is 1.96 multiplied by the standard error. In these examples, we have presented what was originally proposed, with context usually dictating the most suitable metric, which will be guided by the purpose of the plot.

Although these recommendations give a clear steer on the type of visualisations to consider, with some guiding principles on format, users can vary many aspects of plot design. For example, colours and symbols used, axis scales and limits, text formatting, appropriate use of labels, and number of groups being compared at once can all affect interpretation and understanding. Much has been written on these aspects, and we refer readers to the articles by Unwin and Muth,[15 27] as well as lists of key principles for a good visualisation in several publications.[1 28 29]

Strengths and limitations of this work

The predominance of statisticians over other researchers in the consensus group could be deemed a limitation of this work. Statisticians are, however, typically responsible for producing information on harms, such as in tables or visualisations, and thus the implementation of these recommendations. We therefore deemed their inputs and opinions highly relevant to the process. In addition to statisticians, a graphic designer was present across all meetings, and feedback was sought from each continually throughout the project. To ensure breadth of input, we worked with clinicians with experience in clinical trials to seek their feedback on the endorsed plots and to ensure understanding of each plot because they are likely to be the main consumers of such information. This collaboration with clinicians allowed us to incorporate clarifications into the recommendations where necessary. We hope that choosing clinicians who are active trialists will help to assist with dissemination of our findings and help us to increase the likelihood of these plots being used in practice. Patients were not involved in this work because our focus was to identify the best plots to present in scientific journals with a predominantly scientific readership. Our aim was to first provide guidance and tools to the authors of reports of randomised controlled trials. The next step that needs to be addressed is patient feedback. We did not consider use of interactive visualisations in these recommendations because we believe that these are in their own separate domain and require different considerations for appraisal (see Wang et al[30]). Given the multifaceted, complex nature of data for harms and advances in the way readers consume and access journal articles, interactivity could be highly advantageous for future projects.

Several novel visuals were considered for endorsement in this work (eg, the volcano and tendril plot shown in supplement 3 figs A11 and A15), but the appraisals showed their inadequacies and a preference for more traditional plots. Endorsement was given for two less commonly used plots—the survival ratio plot and the plot of the mean cumulative frequency, and we encourage use of such plots with clear explanations to ease interpretation. We particularly encourage use of the mean cumulative function plot as a summary of the overall burden of harm in place of, or in addition to, summaries of time to discontinuation that are often reported as a proxy for harm. Given the scarcity of visualisations for presenting data for harm outcomes for randomised controlled trials, use of any visualisation of these harms is arguably novel, especially for emerging events. Once the use of visualisations for harm outcomes is more common in scientific publications, the desire for more innovative plots might increase.

Although we suggest amendments to existing plots, the purpose of this work was not to develop new plots. However, it was clear that new approaches are needed for some scenarios, particularly when the visualisation of multiple time-to-event outcomes or multiple continuous outcomes is of interest, or when consideration of duration of events is important. Development of new plots will be undertaken in future work and we will seek to update guidelines to reflect any future progress. With a high likelihood of future updates being required, development of a website that can be more readily updated over time without need for new publications is one further thing to explore and has previously been advocated by Chuang-Stein and Xia.[10] This would also serve as a readily available resource for dissemination. The CTSpedia Wiki page created by scientists from industry, academia, and the US Food and Drug Administration goes some way towards this, serving as a repository of potential visualisations, although it provides limited direction on benefits of each plot, cautions of use, and possible inferences to be drawn; it has not been updated since 2014.[1]

Conclusions

Visualisations provide a powerful tool to communicate harms in clinical trials, offering an alternative perspective to the traditional frequency tables. Implementation of the recommendations in this article should improve reporting of harm outcomes in clinical trial manuscripts and enable clearer presentation of harm profiles, and should help to identify potential signals for adverse drug reactions for further monitoring. We endorse each of the plots presented; however, we also highlight the limitations of each plot and provide examples of when their use would be inappropriate. We also caution users to practise care when creating and interpreting each plot. Although the decision tree aids the choice of visualisation, statisticians and clinical trial teams must ultimately decide the most appropriate visualisations for their data and objectives. We recommend trialists continue to examine crude numbers alongside visualisations to fully understand harm profiles. This information should also be reported in supplementary appendices so that readers of trial manuscripts can also appraise this information and so that the data are available to researchers who want to undertake systematic reviews and meta-analyses of harms.[31]

AUTHOR AFFILIATIONS

[1]Imperial Clinical Trials Unit, School of Public Health, Imperial College London, London, UK

[2]Pragmatic Clinical Trials Unit, Centre for Evaluation and Methods, Wolfson Institute of Population Health, Queen Mary University of London, London, UK

[3]Cambridge Clinical Trials Unit, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK

[4]MRC Clinical Trials Unit at University College London, Institute of Clinical Trials and Methodology, London, UK

[5]Exeter Clinical Trials Unit, University of Exeter, Exeter, UK

[6]York Trials Unit, University of York, York, UK

[7]CRUK Cancer Trials Centre, University College London, London, UK

[8]Nuffield Department of Population Health, University of Oxford, Oxford, UK

[9]Liverpool Clinical Trials Centre, University of Liverpool, Liverpool, UK

[10]Centre for Health Care Randomised Trials, University of Aberdeen, Aberdeen, UK

[11]Roche Products, Welwyn Garden City, UK

[12]Division of Anaesthetics, Pain Medicine, and Intensive Care, Department of Surgery and Cancer, Imperial College London and Imperial College Healthcare NHS Trust, London, UK

[13]Imperial College London and Imperial NHS Trust, London, UK

[14]Department of Ophthalmology, School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

[15]Clinical Trials Research Unit, Leeds Institute for Clinical Trials Research, University of Leeds, Leeds, UK

[16]Oxford Clinical Trials Research Unit, Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Sciences, University of Oxford, Oxford, UK

[17]Cancer Research UK Clinical Trials Unit, University of Birmingham, Birmingham, UK

[18]Royal Marsden Clinical Trials Unit, Royal Marsden NHS Foundation Trust, London, UK

[19]Comprehensive Clinical Trials Unit, University College London, London, UK

[20]Bristol Trials Centre, University of Bristol, Bristol, UK

We thank Will Stahl-Timmins (data graphics designer at *The BMJ*) for his inputs and guidance throughout this project and Rachel Evans (NWORTH clinical trial unit, Bangor University) for her contributions over the course of the consensus meetings.

Contributors: RP and VC conceived and designed this project. RP prepared materials for consensus meetings with feedback provided by VC. RP, VC, and SC designed and facilitated the consensus meetings. RP was responsible for undertaking the interviews with clinicians. All authors were involved in discussions and development of the recommendations. RP wrote the first draft of the manuscript and VC provided initial critical revision. All authors critically reviewed, revised, and approved the manuscript for submission. VC provided supervision of this project. RP is guarantor and attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding: RP was funded by a National Institute for Health and Care Research (NIHR) doctoral research fellowship to undertake this work (DRF-2017-10-131). SC is supported by an NIHR advanced fellowship (NIHR300593). This paper presents independent research funded by NIHR. The views expressed are those of the authors and not necessarily those of the NHS, NIHR, or Department of Health and Social Care. Infrastructure support for this research was provided by the Imperial College London NIHR Biomedical Research Centre. TPM was supported by MRC grants MC_UU_00004/07, MC_UU_12023/21 and MC_UU_12023/29. The funders had no role in considering the study design or in the collection, analysis, interpretation of data, writing of the report, or decision to submit the article for publication. The UK Clinical Research Collaboration Registered Clinical Trials Unit Statistics Operations Group supports and endorses this work.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/disclosure-of-interest/ and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: This study was approved by the Imperial Joint Research Compliance Office in February 2019(ethics No 19IC5071).

Data sharing: The datasets used in this analysis are available from GlaxoSmithKline via ClinicalStudyDataRequest.com, but restrictions apply to the availability of these data, which were used under licence for the current study. The synthetic dataset example is available for download in the Stata aedot and aevolcano command packages.

The lead author (RP) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Provenance and peer review: Not commissioned; externally peer reviewed.

Dissemination to participants and related patient and public communities: All parties that participated in the development of these recommendations have been sent a copy of this manuscript and encouraged to share the recommendations within their own research communities. Dissemination to the wider clinical trial community will be undertaken via presentations at relevant conferences and seminars.

1 Duke SP, Bancken F, Crowe B, Soukup M, Botsis T, Forshee R. Seeing is believing: good graphic design principles for medical research. *Stat Med* 2015;34:3040-59. doi:10.1002/sim.6549

2 Tufte ER. *The Visual Display of Quantitative Information.* 2nd ed. Graphics Press, 2001.

3 Edwards IR, Biriell C. Harmonisation in pharmacovigilance. *Drug Saf* 1994;10:93-102. doi:10.2165/00002018-199410020-00001

4 International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). ICH Harmonised Tripartite Guideline. E2A Clinical safety data management: definitions and standards for expedited reporting, 1994.

5 Ioannidis JP, Evans SJ, Gøtzsche PC, et al, CONSORT Group. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med* 2004;141:781-8. doi:10.7326/0003-4819-141-10-200411160-00009

6 Lineberry N, Berlin JA, Mansi B, et al. Recommendations to improve adverse event reporting in clinical trial publications: a joint pharmaceutical industry/journal editor perspective. *BMJ* 2016;355:i5078. doi:10.1136/bmj.i5078

7 Crowe BJ, Xia HA, Berlin JA, et al. Recommendations for safety planning, data collection, evaluation and reporting during drug, biologic and vaccine development: a report of the safety planning, evaluation, and reporting team. *Clin Trials* 2009;6:430-40. doi:10.1177/1740774509344101

8 International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). ICH Topic E9 Statistical Principles for Clinical Trials, 1998.

9 Amit O, Heiberger RM, Lane PW. Graphical approaches to the analysis of safety data from clinical trials. *Pharm Stat* 2008;7:20-35. doi:10.1002/pst.254

10 Chuang-Stein C, Xia HA. The practice of pre-marketing safety assessment in drug development. *J Biopharm Stat* 2013;23:3-25. doi:10.1080/10543406.2013.736805

11 Phillips R, Hazell L, Sauzet O, Cornelius V. Analysis and reporting of adverse events in randomised controlled trials: a review. *BMJ Open* 2019;9:e024537. doi:10.1136/bmjopen-2018-024537

12 Phillips R, Cornelius V. Understanding current practice, identifying barriers and exploring priorities for adverse event analysis in randomised controlled trials: an online, cross-sectional survey of statisticians from academia and industry. *BMJ Open* 2020;10:e036875. doi:10.1136/bmjopen-2020-036875

13 Colopy MW, Gordon R, Ahmad F, Wang WW, Duke SP, Ball G. Statistical Practices of Safety Monitoring: An Industry Survey. *Ther Innov Regul Sci* 2019;53:293-300. doi:10.1177/2168479018779973

14 Gelman A, Pasarica C, Dodhia R. Let's practice what we preach. *Am Stat* 2002;56:121-30. doi:10.1198/000313002317572790

15 Unwin A. Why is data visualization important? what is important in data visualization?*Harvard Data Sci Rev* 2020;2. doi:10.1162/99608f92.8ae4d525

16 Singh S, Loke YK. Drug safety assessment in clinical trials: methodological challenges and opportunities. *Trials* 2012;13:138. doi:10.1186/1745-6215-13-138

17 Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Stat Med* 2004;23:1351-75. doi:10.1002/sim.1761

18 Bradburn MJ, Deeks JJ, Berlin JA, Russell Localio A. Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. *Stat Med* 2007;26:53-77. doi:10.1002/sim.2528

19 AEDOT. *Stata module to produce dot plot for adverse event data.* [program] Boston College Department of Economics, 2020.

20 Thanarajasingam G, Atherton PJ, Novotny PJ, Loprinzi CL, Sloan JA, Grothey A. Longitudinal adverse event assessment in oncology clinical trials: the Toxicity over Time (ToxT) analysis of Alliance trials NCCTG N9741 and 979254. *Lancet Oncol* 2016;17:663-70. doi:10.1016/S1470-2045(16)00038-3

21 Proctor T, Schumacher M. Analysing adverse events by time-to-event models: the CLEOPATRA study. *Pharm Stat* 2016;15:306-14. doi:10.1002/pst.1758

22 Rogers JK, Pocock SJ, McMurray JJV, et al. Analysing recurrent hospitalizations in heart failure: a review of statistical methodology, with application to CHARM-Preserved. *Eur J Heart Fail* 2014;16:33-40. doi:10.1002/ejhf.29

23 Morris TP, Jarvis CI, Cragg W, Phillips PPJ, Choodari-Oskooei B, Sydes MR. Proposals on Kaplan-Meier plots in medical research and a survey of stakeholder views: KMunicate. *BMJ Open* 2019;9:e030215. doi:10.1136/bmjopen-2019-030215

24 Proctor T, Schumacher M. Analysing adverse events by time-to-event models: the CLEOPATRA study. *Pharm Stat* 2016;15:306-14. doi:10.1002/pst.1758

25 Newell J, Kay JW, Aitchison TC. Survival ratio plots with permutation envelopes in survival data problems. *Comput Biol Med* 2006;36:526-41. doi:10.1016/j.compbiomed.2005.03.005

26 Gamble C, Krishan A, Stocken D, et al. Guidelines for the content of statistical analysis plans in clinical trials. *JAMA* 2017;318:2337-43. doi:10.1001/jama.2017.18556

27 Muth LC. *Thoughts & How To's. How to pick more beautiful colors for your data visualizations.* Chartable, 2020. https://blog.datawrapper.de/beautifulcolors

28 Vandemeulebroecke M, Baillie M, Margolskee A, Magnusson B. Effective visual communication for the quantitative scientist. *CPT Pharmacometrics Syst Pharmacol* 2019;8:705-19. doi:10.1002/psp4.12455

29 Vickers AJ, Assel MJ, Sjoberg DD, et al. Guidelines for reporting of figures and tables for clinical research in urology. *Eur Urol* 2020;78:97-109. doi:10.1016/j.eururo.2020.04.048

30 Wang W, Revis R, Nilsson M, et al. Clinical trial drug safety assessment with interactive visual analytics. *Stat Biopharm Res* 2020:1-12. doi:10.1080/19466315.2020.1736142

31 Cornelius V, Cro S, Phillips R. Advantages of visualisations to evaluate and communicate adverse event information in randomised controlled trials. *Trials* 2020;21:1028. doi:10.1186/s13063-020-04903-0

Web appendix: Supplement 1: methodological details

Web appendix: Supplement 2: recommended plots

Web appendix: Supplement 3: Visualisations considered but not recommended

Web appendix: Supplement 4: Example R code and dataset to create the dot plot

Web appendix: Supplement 5: Table and figures summarising initial appraisals of all plots by outcome type

Web appendix: Supplement 6: Tables summarising Mentimeter votes to decide which plots to take forward and amendments

Web appendix: Supplement 7: Free text comments accompanying initial appraisals of recommended plots

Supplement 1: methodological details

EXAMPLE DATASETS

We produced each of the candidate plots using data from four parallel arm pharmacological RCTs, obtained via the ClinicalStudyDataRequest.com initiative from GlaxoSmithKline (GSK). The first was a two-arm (1:1 allocation) study that evaluated the efficacy of mepolizumab compared to placebo in patients with severe eosinophilic asthma (n=135) (ClinicalTrials.gov number: NCT01691508). The second study investigated mepolizumab in patients with severe uncontrolled refractory asthma comparing two doses of mepolizumab to placebo (1:1:1) (n=576) (ClinicalTrials.gov number: NCT01691521). The third study was a two-arm (2:1) trial examining the efficacy, safety and tolerability of paroxetine compared to placebo in adolescents with unipolar major depression (n=286). The fourth was a two-arm (1:1) trial examining the efficacy and tolerability of paroxetine compared to placebo in paediatric major depression (n=206).[1-4] In addition, a synthetic dataset was created based on a RCT of a novel active treatment for eczema compared to placebo (1:1) in adolescents unresponsive to standard care (n=61) (the synthetic dataset is available for download in the Stata aedot and aevolcano command packages).[5 6]

METHODS

Review

A methodology review performed in March 2018 and updated up until October 2019 identified statistical methods specifically developed to analyse harm outcomes, including the use of visualisations.[7] The review identified over 20 unique methods to visually summarise harm data, including binary AEs and continuous laboratory (e.g. blood tests, culture data) and vital signs (e.g. temperature, blood pressure, electrocardiograms) data.[8-15] These identified visualisations were taken forward for evaluation at the consensus meeting. In addition, alternative visualisations that could be adapted to the harm setting were also considered.[16]

Supplement 1: methodological details

The available graphics were categorised according to the type and number of outcome they support and are presented according to these categories in the following sections. Type of outcomes considered included: binary harm outcomes which includes events such as occurrence of a headache or experiencing nausea, count outcomes i.e. the number of occurrences of an event which could include number of headaches experienced over follow-up, time-to-event outcomes which could include time from treatment exposure to headache and continuous outcomes such as individual results from a blood count. Plots considered were suitable for displaying either single outcomes or multiple simultaneously.

Recommendation development

Consensus meeting

In February 2020, the lead organisers (RP, VC and SC) sent emails to members of the UKCRC CTU Statisticians' Operations Group and personal contacts in academia and industry with a known interest in visualisations, and an advert was placed on the PSI (Statisticians in the Pharmaceutical Industry) visualisation special interest group (SIG) homepage seeking researchers with applied experience of analysing trial data. The emails invited recipients to participate in a consensus meeting to take place over the course of three half-day virtual sessions in July 2020.

Twenty-seven participants were invited to attend in line with the CONSORT group executive recommendations to limit meetings to no more than 30 participants.[17] Twenty-three participants contributed to at least one of the sessions over the course of the three days and are listed in the authorship or acknowledgments. This included 20 statisticians from 15 UKCRC registered CTUs, a health economist based at an academic population health department, one industry statistician, and a data graphics designer who sits on the multimedia team at the BMJ.

Supplement 1: methodological details

A week in advance of the first meeting lead organisers shared with participants the graphics proposed for visually summarising harm data identified in the methodological review, graphics proposed from other settings that could be adapted to the harm setting and any suggested initial adaptations, along with a proposed framework for appraisal. This also included a call for suggestions of any plots that may have been inadvertently omitted.

A draft framework for appraisal was developed in advance taking into consideration work from Ballarini et al. who proposed a framework to assess the properties of graphics for subgroup analysis, principles for producing effective visualisations proposed by Gordon and Finch, and discussions amongst lead organisers regarding the important components to communicate when analysing harm outcomes.[18 19] In the first meeting, the draft criteria to appraise each of the graphics were discussed amongst participants and refined based on feedback and group endorsement. The final criteria are included in table A1. Assessment criteria comprised of eight items related to plot content and presentation including: whether the plot clearly displays an effect size for each event; whether it clearly displays a robust measure of uncertainty; and whether it requires supplementary data presentations. Each item was scored on a scale of 1 to 5. Lower scores indicated negative responses such as ‘very unclear’ or ‘very difficult’ or ‘strongly disagree’. Two further items related to suitability for use in journal articles or interim analysis reports and whether the plot was suitable for explanatory or exploratory analysis. Exploratory analysis was defined as visualisations suited to data exploration to help identify potential signals for ADRs and explanatory analysis was defined as visualisation suited to communicate a message about the data. Participants were also asked to rank each plot in order of preference in relation to other plots within the same category.

Supplement 1: methodological details

Over the course of the first two meetings, each of the graphics and a summary of its main elements were presented in turn by category and discussed. Participants were encouraged to use the audio and the chat function to: raise any queries they had regarding each plot; highlight what they liked or disliked about it; consider in which research contexts they thought it might be useful; and raise any potential problems or opportunities for causing confusion. Participants completed their appraisals for each graphic and were encouraged to include free text comments if they endorsed any of the specific recommendations or adaptations discussed.

Appraisals were returned to the lead organisers following each meeting and results were summarised and shared with participants in advance of the next meeting to inform discussions (summary scores are presented in supplement 5 tables A.2-A.7 and figures A.29-A.34). After examining the initial appraisal results, participants were encouraged to champion low-scoring plots if they felt strongly that they were under-scored and asked to consider where possible adaptations might be needed. Once discussions were concluded, participants voted on whether to take plots forward for further discussion around recommendations for use and refinements. Results of these votes were presented back to the group in real time. If a plot received at least 60% of the available votes, we considered the plot to be endorsed. Scores marginally below this threshold (50-60%) were revisited for further discussions and votes retaken until a consensus could be reached (results summarised in supplement 6 tables A.8-A.14).

For each of the endorsed plots, we summarised the discussions and free text comments from the appraisal sheets and presented them back to the group (summaries of these comments are included supplement 7). Participants were given the opportunity to raise any other points they felt were important but had been omitted. These included comments about potential adaptations, where we would recommend using each plot and any cautions or limitations that should be included within the recommendations. We used Mentimeter to record endorsement for each adaptation and finalise the appearance of the endorsed plots and accompanying recommendations.[20] Endorsed plots with incorporated adaptations are presented in the following.

One-to-one interviews with clinicians

In August 2020, two clinical collaborators who are experienced clinical trialists participated in one-to-one semi-structured interviews with one of the lead organisers (RP) lasting approximately one hour. During the interviews, work to date was outlined and feedback on the consensus group's recommendations was sought. This was an opportunity to gather insights on clinicians understanding of the utility and interpretation of each plot, which could then be used to structure the explanatory information provided in the recommendations. Topics covered included:

a. Opinions on the finalised plots, including the merits of each, and whether they were likely to use/endorse these plots in practice;
b. Whether they thought any of the plots were unclear and/or required further explanation that could be incorporated into the recommendations;
c. Whether they thought any modifications were required to any of the plots.

Questions asked were open-ended to allow for detailed responses and comments. We also encouraged clinical participants to raise any other comments they had that were not prompted from the topic areas and invited them to provide written feedback following the meeting if they wished to. Pertinent comments raised were incorporated into the recommendations. Both clinicians are listed as co-authors (AG and NV).

Supplement 1: methodological details

Table A.1: Framework for assessing the properties of graphical displays

Item	Criterion for appraisal	Response options
1	Effect size - Does it clearly display an effect size for events?	1: no/very unclear, 2: unclear, 3: unsure, 4: clear, 5: yes/very clear
2	Treatment effect - Does it clearly display the direction of the treatment effect?	1: no/very unclear, 2: unclear, 3: unsure, 4: clear, 5: yes/very clear
3	Uncertainty – Does it clearly display a robust measure of uncertainty such as CI or SEs?	1: no/very unclear, 2: unclear, 3: unsure, 4: clear, 5: yes/very clear
4	Does it require supplementary data presentations? i.e. Does it stand-alone or does it need additional data presented alongside it?	1: yes/extremely likely, 2: likely, 3: neutral, 4: unlikely, 5: extremely unlikely/stand-alone
5	Can you understand the plot without a detailed explanation?	1: very difficult, 2: difficult, 3: neutral, 4: easy, 5: very easy
6	Do you think non-statistical colleagues i.e. clinicians can understand the plot without a detailed explanation?	1: very difficult, 2: difficult, 3: neutral, 4: easy, 5: very easy
7	How adaptable is the plot for multi-arms/adaptive trials?	1: very difficult, 2: difficult, 3: neutral, 4: easy, 5: very easy
8	Are there limitations around the number of events displayed?	1: yes/extremely limited, 2: very limited, 3: moderately limited, 4: slightly limited, 5: not at all/unlimited
	Total for items 1-7*	
9	Is it suitable for inclusion in a:	
i	Journal article	1: strongly disagree, 2: disagree, 3: neutral, 4: agree, 5: strongly agree
ii	Final study report	1: strongly disagree, 2: disagree, 3: neutral, 4: agree, 5: strongly agree
iii	Interim analysis report	1: strongly disagree, 2: disagree, 3: neutral, 4: agree, 5: strongly agree
10	Is it best suited to[†]:	
i	Exploratory analysis	1: strongly disagree, 2: disagree, 3: neutral, 4: agree, 5: strongly agree
ii	Explanatory analysis	1: strongly disagree, 2: disagree, 3: neutral, 4: agree, 5: strongly agree
11	Ranking	1 - most preferred through to least preferred within category
12	Comments: Please indicate if you support any of the amendments proposed for this plot or any other comments on this plot that are not captured elsewhere	Please provide details of the amendment in case of multiple suggested amendments

Abbreviations: CI - confidence interval; SE – standard error

*Total score is a sum of scores assigned to questions 1 through to 7. Scores to question eight were not included, whilst we thought this was important for consideration, we did not wish to disadvantage plots that could only present a limited number of events, as this might be through design and in fact in some settings is likely to be an advantage. The group discussed this point before the decision was made.

[†] Exploratory analysis was defined as visualisations suited to data exploration to help identify potential signals for adverse (drug) reactions (A(D)Rs) and explanatory analysis was defined as visualisation suited to communicate a message about the data.

Supplement 1: methodological details

References

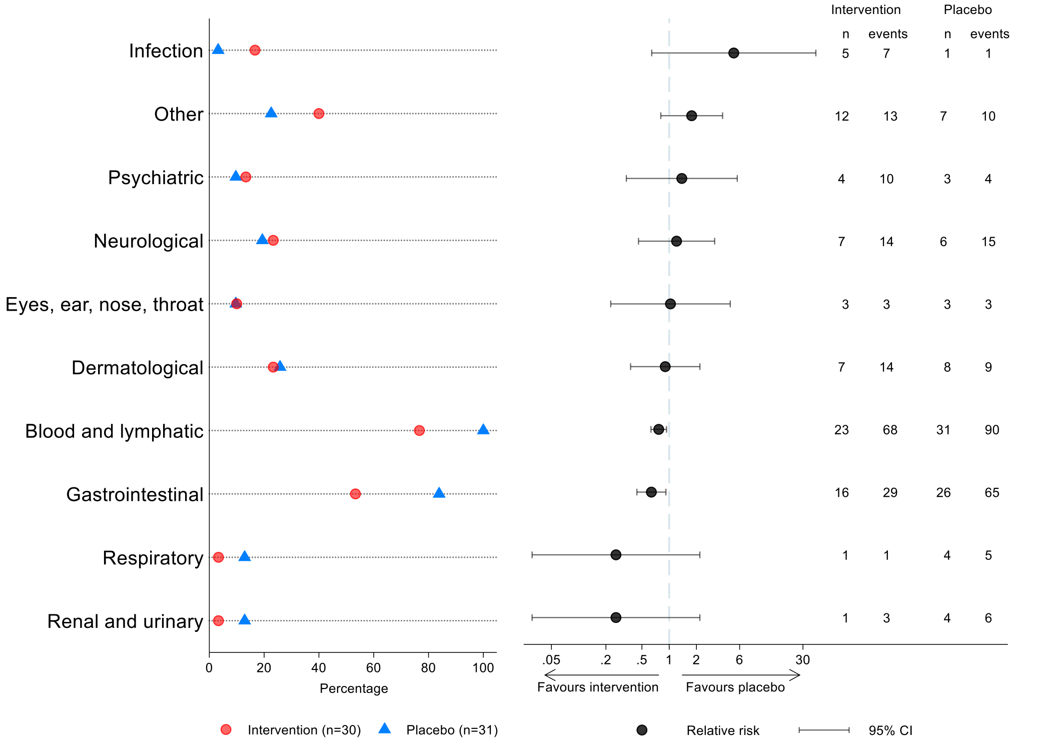
1. Bel EH, Wenzel SE, Thompson PJ, et al. Oral Glucocorticoid-Sparing Effect of Mepolizumab in Eosinophilic Asthma. *New England Journal of Medicine* 2014;371(13):1189-97. doi: 10.1056/NEJMoa1403291
2. Berard R, Fong R, Carpenter DJ, et al. An international, multicenter, placebo-controlled trial of paroxetine in adolescents with major depressive disorder. *J Child Adolesc Psychopharmacol* 2006;16(1-2):59-75. doi: 10.1089/cap.2006.16.59 [published Online First: 2006/03/24]
3. Ortega HG, Liu MC, Pavord ID, et al. Mepolizumab Treatment in Patients with Severe Eosinophilic Asthma. *New England Journal of Medicine* 2014;371(13):1198-207. doi: 10.1056/NEJMoa1403290
4. Emslie GJ, Wagner KD, Kutcher S, et al. Paroxetine Treatment in Children and Adolescents With Major Depressive Disorder: A Randomized, Multicenter, Double-Blind, Placebo-Controlled Trial. *Journal of the American Academy of Child & Adolescent Psychiatry* 2006;45(6):709-19. doi: https://doi.org/10.1097/01.chi.0000214189.73240.63
5. AEDOT: Stata module to produce dot plot for adverse event data [program]: Boston College Department of Economics, 2020.
6. AEVOLCANO: Stata module to produce volcano plot for adverse event data [program]: Boston College Department of Economics, 2020.
7. Phillips R, Sauzet O, Cornelius V. Statistical methods for the analysis of adverse event data in randomised controlled trials: a scoping review and taxonomy. *BMC Medical Research Methodology* 2020;20(1):288. doi: 10.1186/s12874-020-01167-9
8. Zink RC, Wolfinger RD, Mann G. Summarizing the incidence of adverse events using volcano plots and time intervals. *Clinical Trials* 2013;10(3):398-406.
9. Amit O, Heiberger RM, Lane PW. Graphical approaches to the analysis of safety data from clinical trials. *Pharmaceutical Statistics* 2008;7(1):20-35.
10. Chuang-Stein C, Xia HA. The practice of pre-marketing safety assessment in drug development. *Journal of Biopharmaceutical Statistics* 2013;23(1):3-25. doi: 10.1080/10543406.2013.736805
11. Chuang-Stein C, Le V, Chen W. Recent Advancements in the Analysis and Presentation of Safety Data. *Drug Information Journal* 2001;35(2):377-97. doi: 10.1177/009286150103500207
12. Southworth H. Detecting outliers in multivariate laboratory data. *Journal of Biopharmaceutical Statistics* 2008;18(6):1178-83.
13. Trost DC, Freston JW. Vector Analysis to Detect Hepatotoxicity Signals in Drug Development. *Therapeutic Innovation & Regulatory Science* 2008;42(1):27-34. doi: 10.1177/009286150804200106
14. Karpefors M, Weatherall J. The Tendril Plot—a novel visual summary of the incidence, significance and temporal aspects of adverse events in clinical trials. *Journal of the American Medical Informatics Association* 2018;25(8):1069-73. doi: 10.1093/jamia/ocy016
15. Zink RC, Marchenko O, Sanchez-Kam M, et al. Sources of Safety Data and Statistical Strategies for Design and Analysis:Clinical Trials. *Therapeutic Innovation & Regulatory Science* 2018;52(2):141-58. doi: 10.1177/2168479017738980
16. Phillips RC, VR.; Cro, S.; Sauzet, O. The use of visual analytics for clinical trial safety outcomes: a methodological review. *Trials Meeting abstracts from the 5th International Clinical Trials Methodology Conference (ICTMC 2019)* 2019;20(Supplement 1)
17. Moher D, Schulz KF, Simera I, et al. Guidance for Developers of Health Research Reporting Guidelines. *PLOS Medicine* 2010;7(2):e1000217. doi: 10.1371/journal.pmed.1000217
18. Ballarini NM, Chiu Y-D, König F, et al. A critical review of graphics for subgroup analyses in clinical trials. *Pharmaceutical Statistics* 2020 doi: 10.1002/pst.2012 [published Online First: 25 March 2020]
19. Gordon I, Finch S. Statistician Heal Thyself: Have We Lost the Plot? *Journal of Computational and Graphical Statistics* 2015;24(4):1210-29. doi: 10.1080/10618600.2014.989324

Supplement 1: methodological details

20. Mentimeter [Available from: https://www.mentimeter.com/.

Supplement 2: recommended plots

Figure A.1: Dot plot of events - data taken from the two-arm example dataset with 1:1 allocation ratio

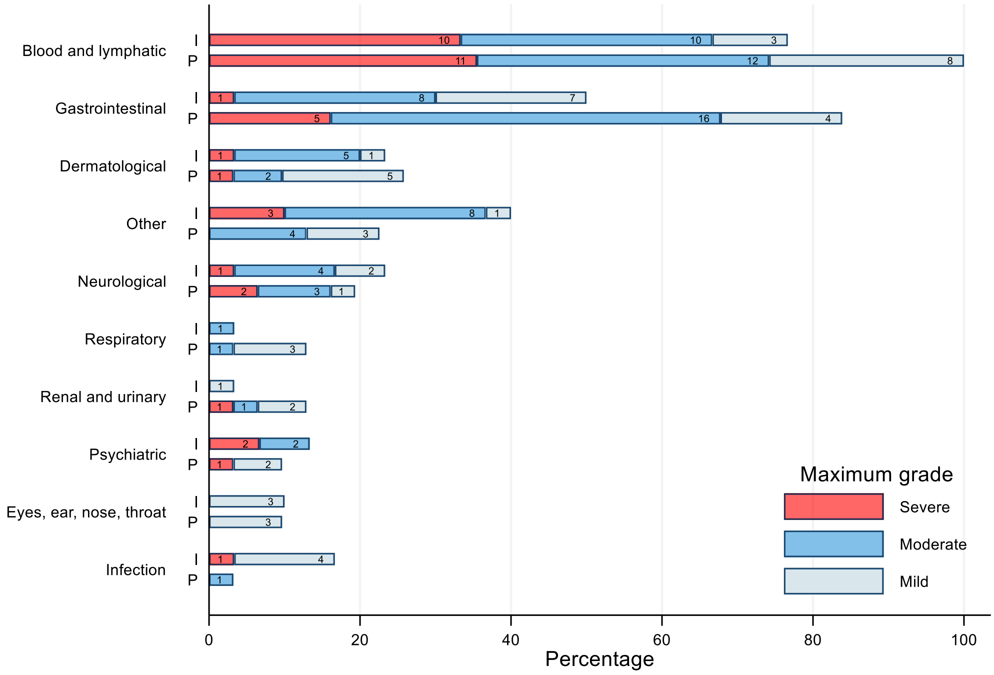


	Intervention		Placebo	
	n	events	n	events
Infection	5	7	1	1
Other	12	13	7	10
Psychiatric	4	10	3	4
Neurological	7	14	6	15
Eyes, ear, nose, throat	3	3	3	3
Dermatological	7	14	8	9
Blood and lymphatic	23	68	31	90
Gastrointestinal	16	29	26	65
Respiratory	1	1	4	5
Renal and urinary	1	3	4	6

Legend: Dot Plot for emerging harm outcomes between two treatment groups for the simulated dataset. The left panel of the figure displays the percentage of participants experiencing an event (labelled on the y-axis) in the intervention group with a red circle and placebo group with a blue triangle. The central panel displays the relative risk and corresponding 95% confidence interval on the log10 scale and a line to show the value of no difference (for relative risks, this is 1). The right panel displays the ‘number of participants experiencing the event at least once’ (n) and ‘the number of events’ (events) (accounting for recurrent events within participants) by treatment group. The dot plot provides a comprehensive visual representation of the entire harm profile.

Supplement 2: recommended plots

Figure A.2: Horizontal stacked bar chart of events by maximum severity – data taken from the two-arm example dataset with 1:1 allocation ratio

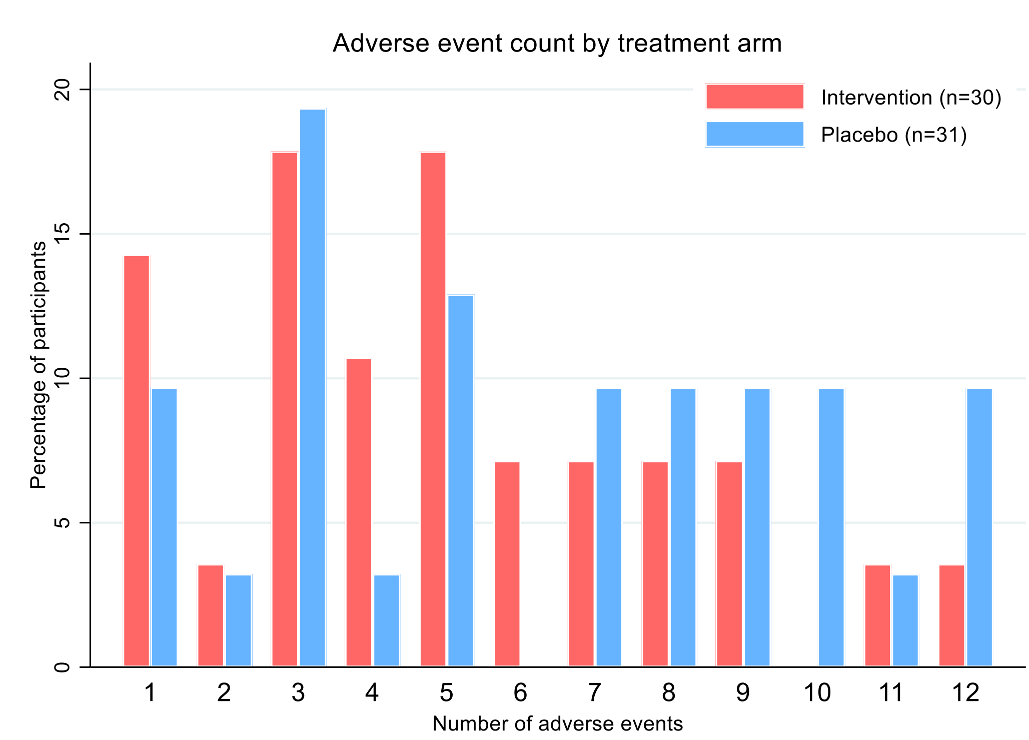


I: Intervention (n=30), P:Placebo (n=31)
Bars are labelled with the corresponding number of participants

Legend: Horizontal stacked bar chart for emerging harm outcomes by maximum severity and treatment group for the simulated dataset. Total bar height represents the proportion of participants with that event at least once and each bar is split into segments to indicate numbers by severity grading. Bar segments are labelled with the corresponding number of participants. The stacked bar chart used in this way is helpful when it is important to present information on the severity of multiple events.

Supplement 2: recommended plots

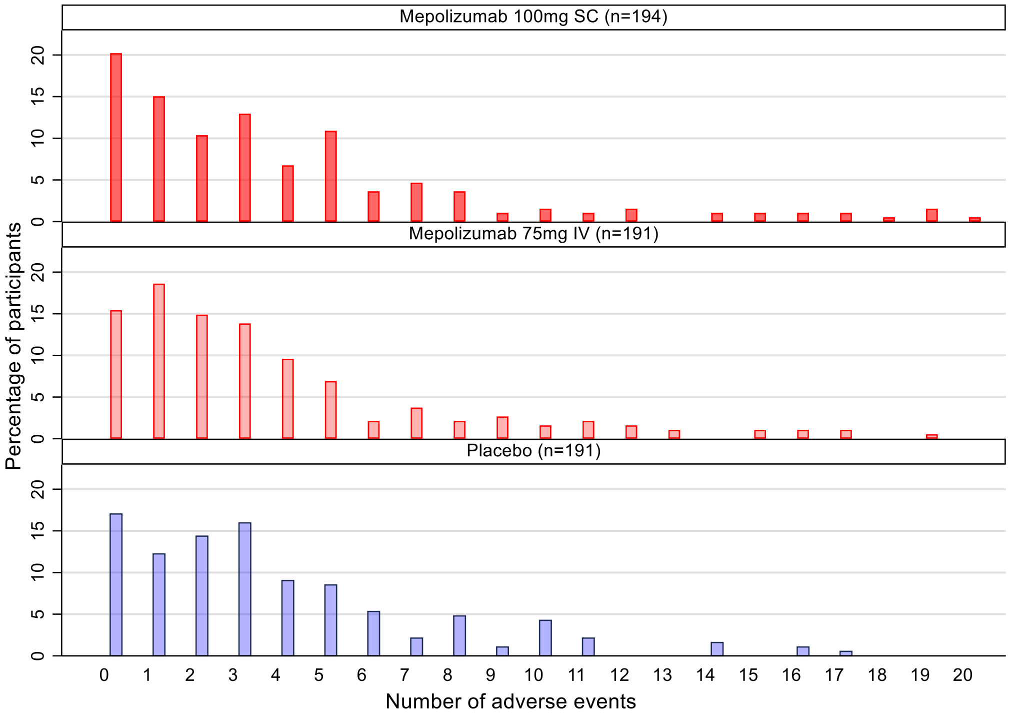
Figure A.3a: Bar chart of event counts – data taken from the two-arm example dataset with 1:1 allocation ratio



Legend: Bar chart of counts of harm outcomes by treatment group for simulated dataset. Each bar represents the proportion of participants with 0, 1, 2 etc. events for each treatment group. This plot groups all adverse events together. Alternatively, it can be used to summarise this information for specific events of interest. Using the bar chart to present this information can help highlight between group differences in the burden of harm experienced by participants.

Supplement 2: recommended plots

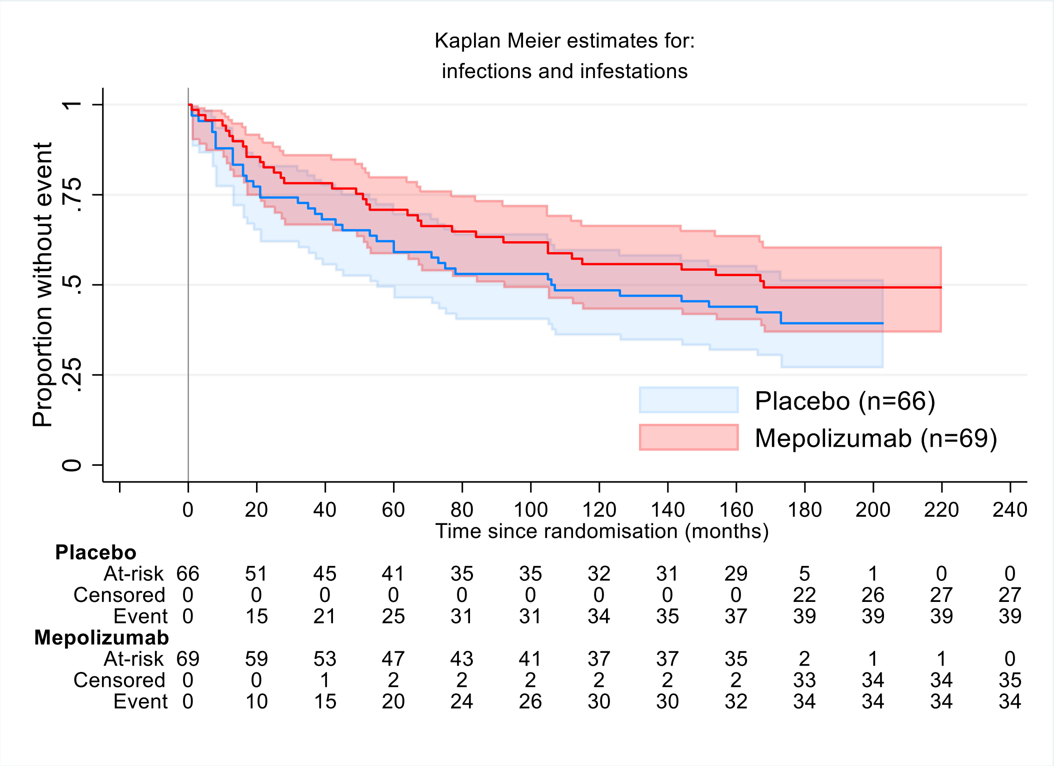
Figure A.3b: Bar chart of event counts – data taken from the three-arm Mepolizumab dataset with 1:1 allocation ratio



Legend: Bar chart of counts of harm outcomes by treatment group (when > 2 treatment groups). Each bar represents the proportion of participants with 0, 1, 2 etc. events for each treatment group. We recommend separate stacked plots like this for trials with more than two treatment group. Using the bar chart to present this information can help highlight between group differences in the burden of harm experienced by participants.

Supplement 2: recommended plots

Figure A.4: Kaplan–Meier plot for an event of interest – data taken from the two-arm Mepolizumab dataset with 1:1 allocation ratio

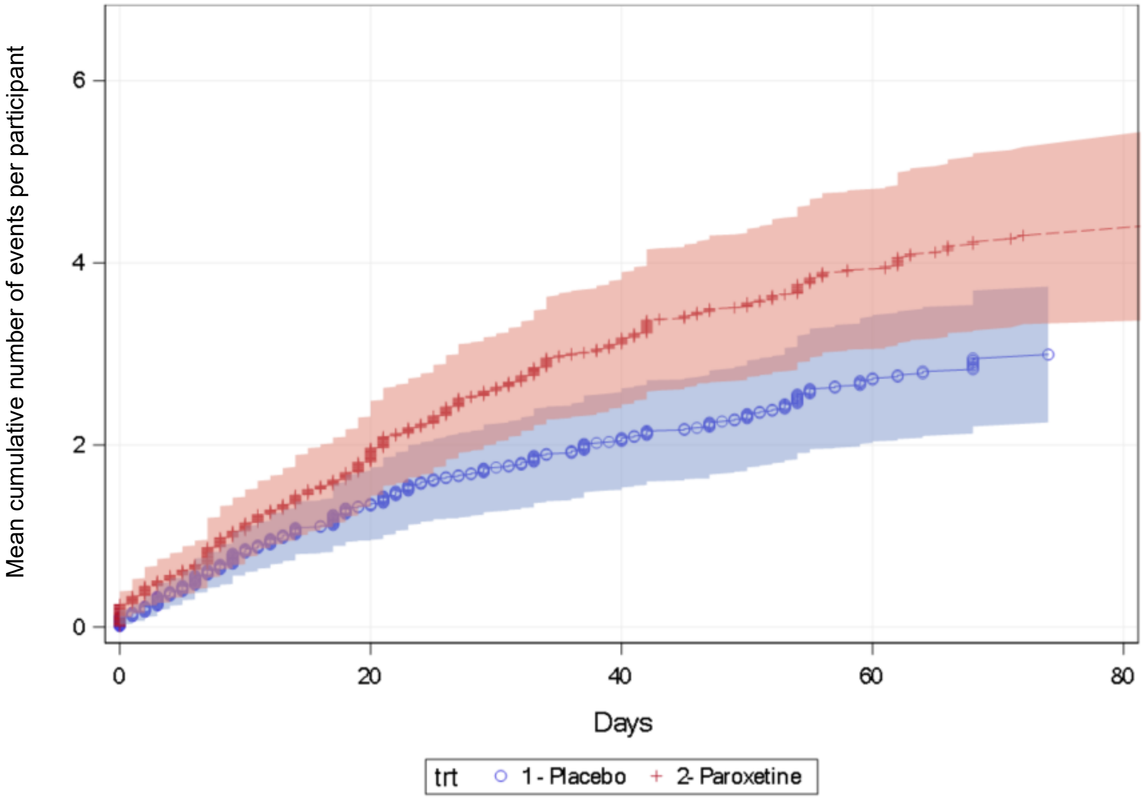


Placebo													
At-risk	66	51	45	41	35	35	32	31	29	5	1	0	0
Censored	0	0	0	0	0	0	0	0	0	22	26	27	27
Event	0	15	21	25	31	31	34	35	37	39	39	39	39
Mepolizumab													
At-risk	69	59	53	47	43	41	37	37	35	2	1	1	0
Censored	0	0	1	2	2	2	2	2	2	33	34	34	35
Event	0	10	15	20	24	26	30	30	32	34	34	34	34

Legend: Kaplan-Meier plot with an extended at risk table for specific harm outcome of interest by treatment group for the two-arm Mepolizumab study. Plots the survival estimates by treatment group where each line indicates the cumulative proportion of participants remaining event free over time by treatment group with 95% confidence intervals for each group separately. The extended at risk table includes information on the number of participants that remain 'at risk', the cumulative number that have been censored and the cumulative number that have experienced an event at discrete time points. In the harm setting, Kaplan-Meier plots can be used to present information for specific events of interest as a useful way to detect a potential disproportionality between treatment groups, which is useful when trying to identify signals for adverse (drug) reactions (A(D)Rs).

Supplement 2: recommended plots

Figure A.5: Mean cumulative function plot for all events – data taken from the two-arm Paroxetine dataset with 1:1 allocation ratio

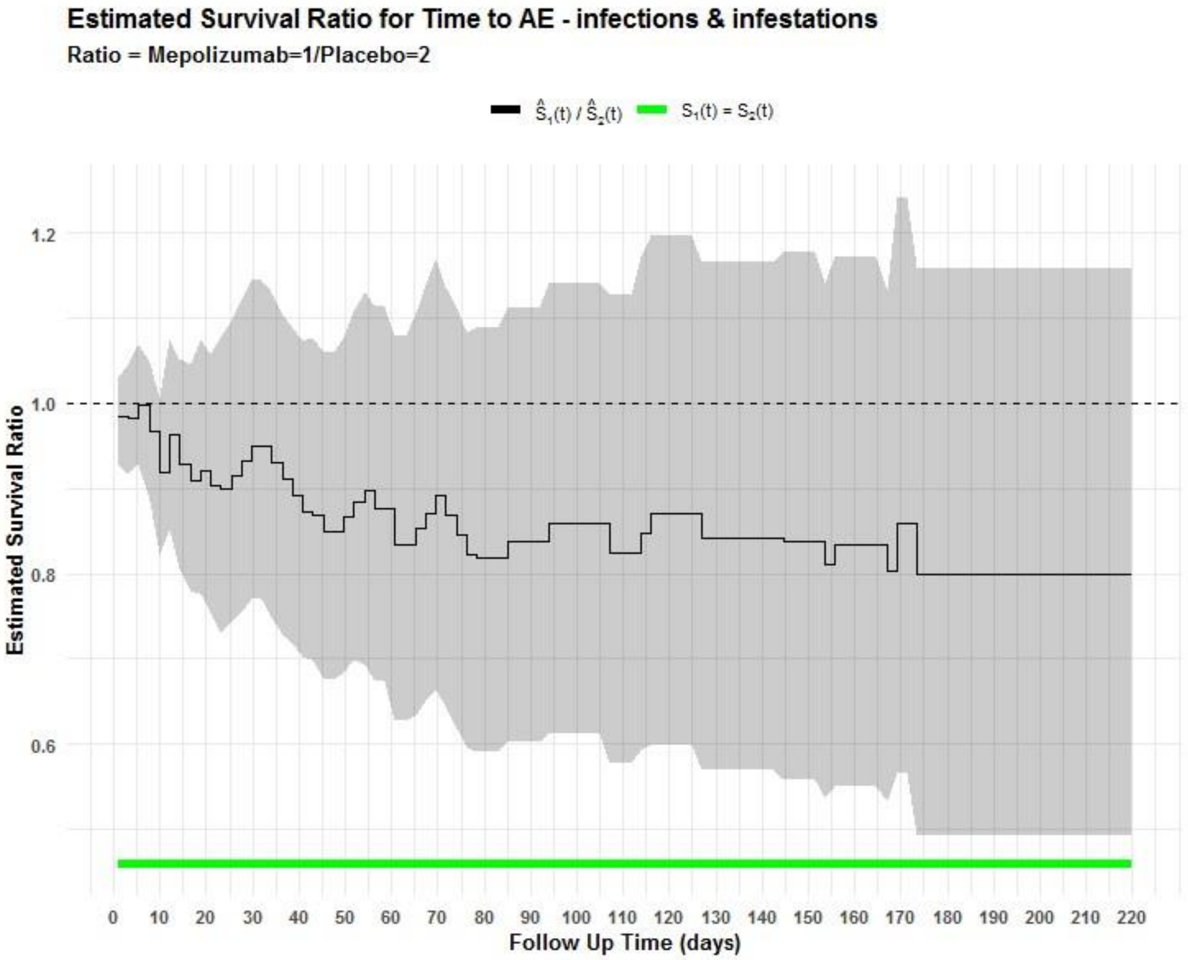


Placebo					
At risk	102	98	91	62	12
Paroxetine					
At risk	101	95	82	60	13

Legend: Mean cumulative function plot for harm outcomes by treatment group for the Paroxetine study with 1:1 treatment allocation. Plots the mean number of events per participant over time by treatment group and includes 95% confidence intervals within groups. The risk table includes information on the number of participants that remain ‘at risk’ at discrete time points throughout the study. In the harm setting, MCF plots can be used to demonstrate a comparison of the burden of experiencing ‘any event’ or the recurrence of events of special interest.

Supplement 2: recommended plots

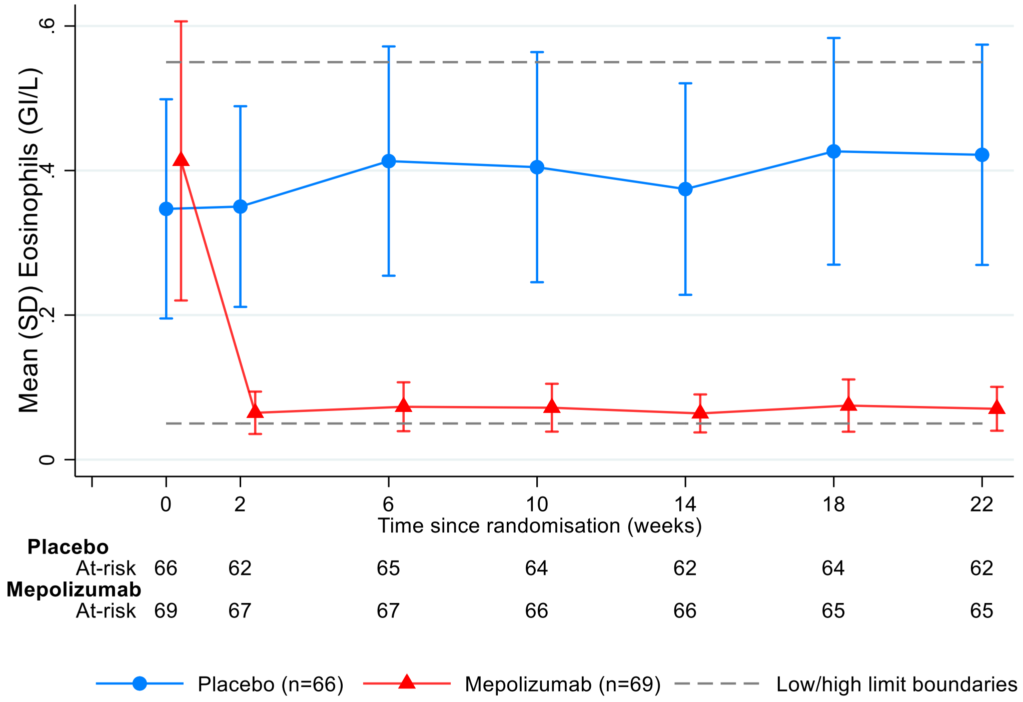
Figure A.6: Survival ratio plot for event of interest – data taken from the two-arm Mepolizumab dataset with 1:1 allocation ratio



Legend: Survival ratio plot for specific harms of interest for two-arm Mepolizumab study. Plots the ratio of survival estimates (solid black line) with the 95% pointwise confidence bands (grey shaded area), where the dashed line at y=1 represents the line of no difference. Departures from unity are indicated using the horizontal band at the bottom of the plot which is green when confidence band includes 1 and red when excludes 1. In the harm setting, the survival ratio plot would be suitable for signal detection analysis across the body of emerging events, as it provides a between group comparison that can be used to detect departures from unity and help identify the time that such divergences occur, which can help detect potential signals for ADRs.

Supplement 2: recommended plots

Figure A.7: Line graph of summary statistic over time by treatment arm for a continuous harm outcome of interest – data taken from the two-arm Mepolizumab dataset with 1:1 allocation ratio

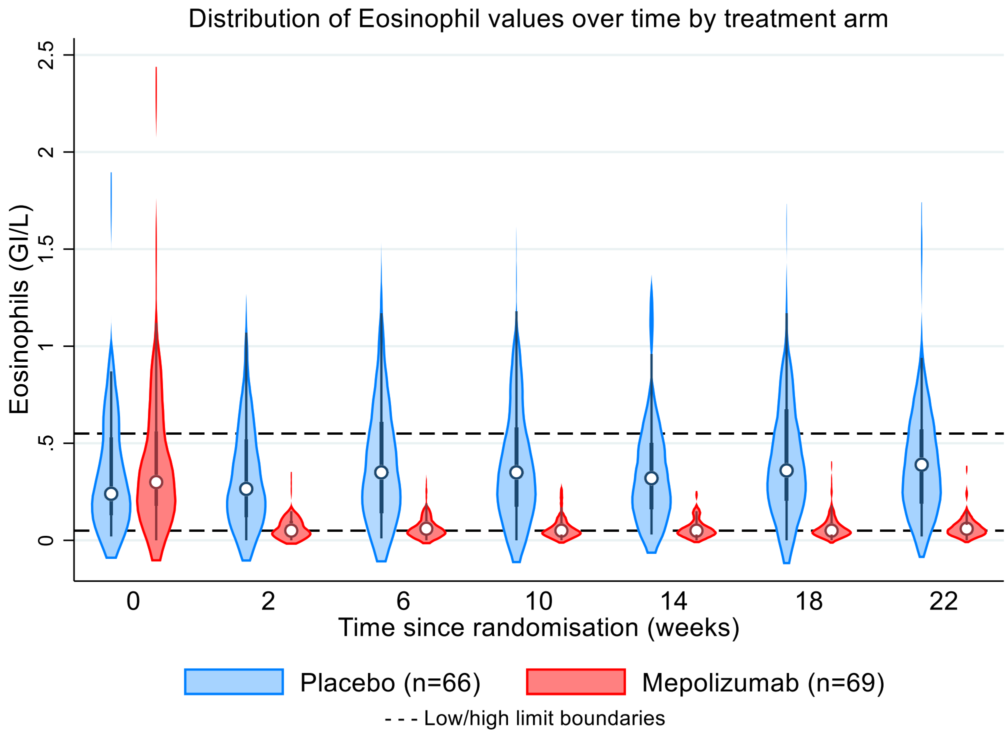


Placebo							
At-risk	66	62	65	64	62	64	62
Mepolizumab							
At-risk	69	67	67	66	66	65	65

Legend: Line graph with risk table for specific continuous outcome of interest by treatment group over time. The markers display an appropriate summary statistic (in this example the mean) and the vertical lines indicate a measure of variability (in this example the standard deviation) of raw values at each discrete point, connected with a line for each treatment group. This plot can be used to describe continuous harm outcomes of interest over time and can help identify shifts in distributions between treatment groups.

Supplement 2: recommended plots

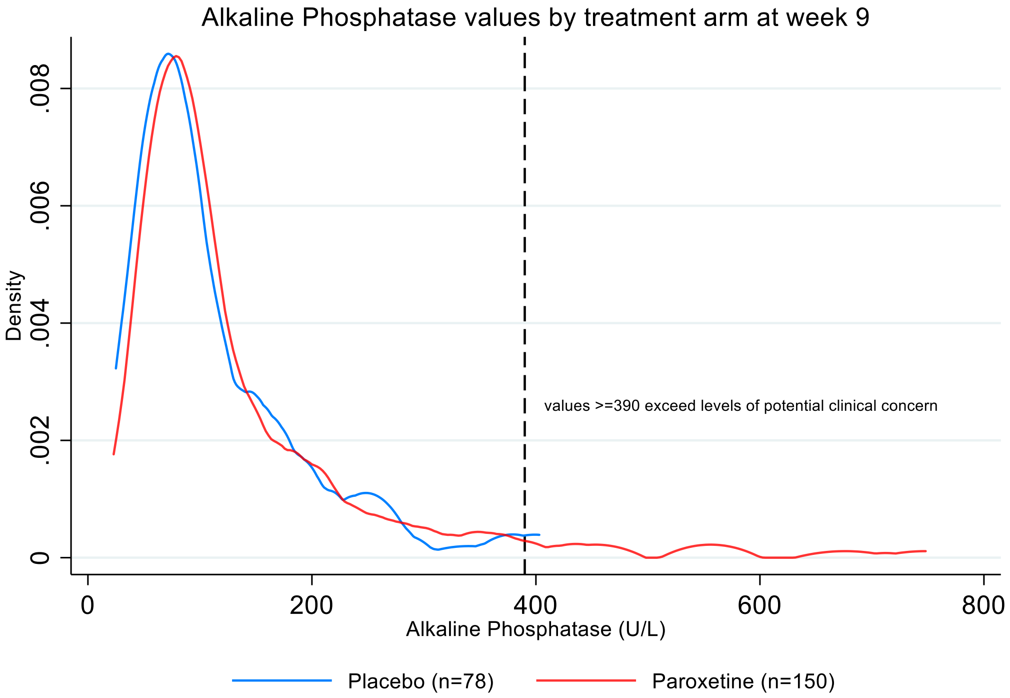
Figure A.8: Violin plot summarising the distribution of a continuous harm outcome of interest over time – data taken from the two-arm Mepolizumab dataset with 1:1 allocation ratio



Legend: Violin plot for specific continuous event of interest by treatment group over time. The hollow circle markers indicate the median, the boxes indicate the inter-quartile range and the lines extend to minimum and maximum points, overlaid with kernel density plots. The violin plot is a useful alternative to the line graph when presenting a continuous outcome that is far from a normal distribution and/or the user is interested in exploring the distribution. It can also help identify outliers and/or identify participants who are persistently showing values of concern.

Supplement 2: recommended plots

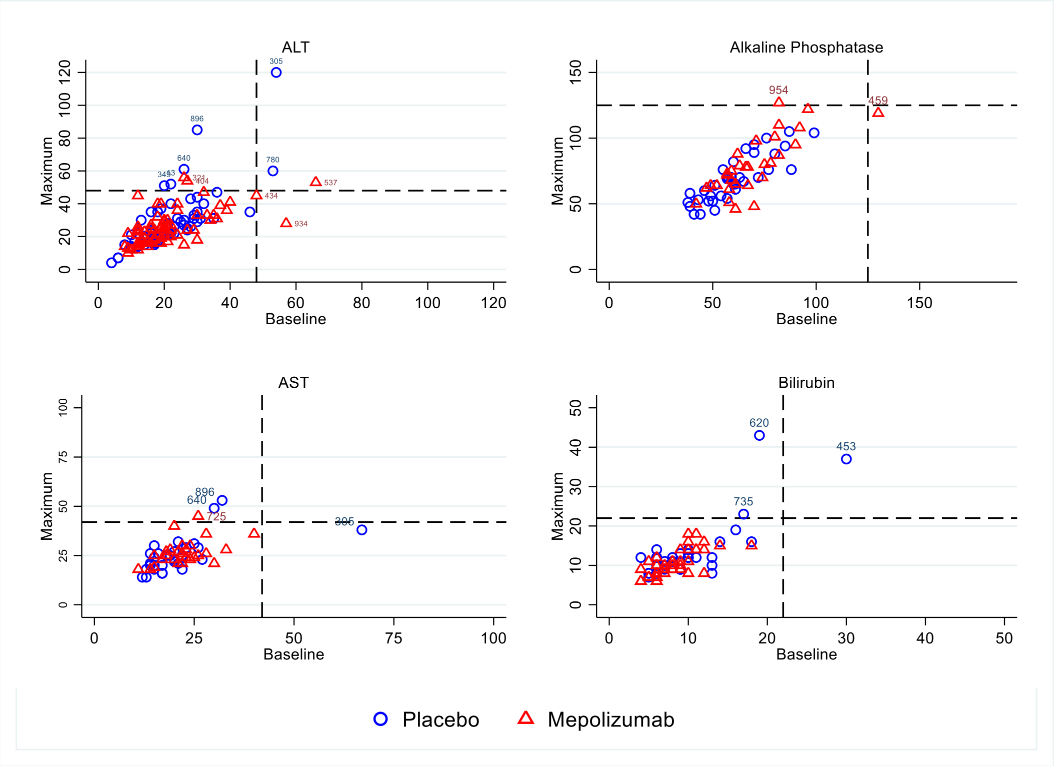
Figure A.9: Kernel density plot for a continuous harm outcome of interest - data taken from the two-arm Paroxetine study with 2:1 allocation ratio



Legend: Kernel density plot for a specific continuous outcome of interest by treatment group, at a single time point, with a reference line to indicate values above which are of clinical concern. It can be helpful to identify shifts in distributions between treatment groups.

Supplement 2: recommended plots

Figure A.10: Scatterplot matrix for continuous harm outcomes – data taken from the two-arm Mepolizumab dataset with 1:1 allocation ratio



Legend: Scatterplot matrix for multiple continuous harm outcomes by treatment group. Plots each participant's baseline value against their maximum on-treatment value. The dashed lines represent the boundary between normal and abnormal thresholds. Outlying observations are labelled with participant identification numbers. This plot can be used in an exploratory setting to identify any outlying observations and to help identify any patterns within participants. In this example where a higher threshold is worse, participants of most concern would be in the top left quadrant (i.e. participants' baseline values were normal and are now abnormal) and the participants who have improved would be in the bottom right (i.e. participants' baseline values were abnormal and are now normal).

Supplement 3 - Visualisations considered but not recommended

Figure A.11: Volcano plot

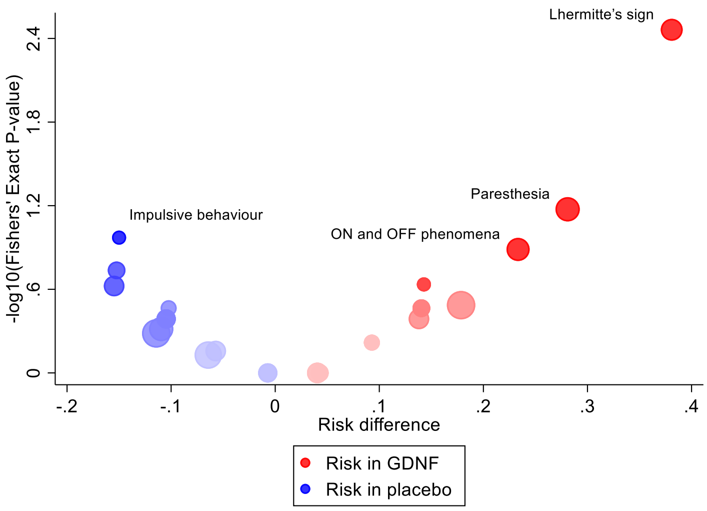


Figure description: Each bubble/circle represents a distinct event. Bubble size/area is proportional to the total number of events across treatment arms. The x-axis indicates the size of the treatment effect. The colour of the bubbles is used to indicate the direction of the treatment effect. The y-axis is used to display log transformed p-values. The colour saturation of each bubble corresponds to the size of the p-value. Under the null hypothesis we would expect to see a U-shape curve with a random scatter of events around the null value, in this case a risk-difference of 0. *Original plot first proposed in: Zink RC, Wolfinger RD and Mann G. Summarizing the incidence of adverse events using volcano plots and time intervals. Clinical Trials 2013; 10: 398-406. Data taken from: Whone A, Luz M, Boca M, et al. Randomized trial of intermittent intraputamenal glial cell line-derived neurotrophic factor in Parkinson's disease. Brain 2019; 142: 512-525*

Adaptions considered: The x-axis could be used to display different metrics e.g. risk difference, odds ratio, incident rate ratios. The p-values displayed on the y-axis can be based on any statistical test and can incorporate a multiple test correction. Colour saturation could be used to reflect an alternative to p-value size such as the average severity rating – this has been suggested by multiple people when demonstrating this plot

Supplement 3 - Visualisations considered but not recommended

Figure A.12: Alternative volcano 1 proposed by BMJ graphic designer (WST)

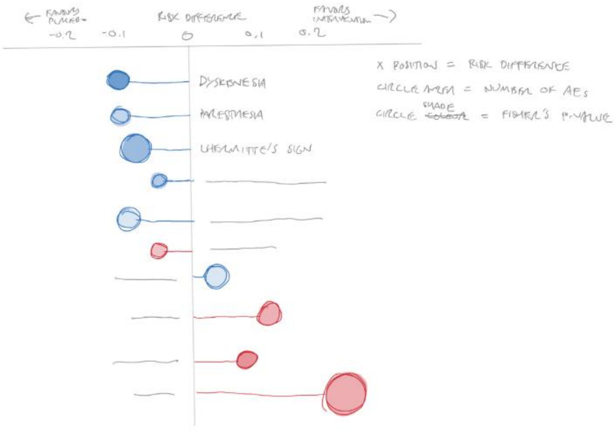


Figure description: Displays the risk difference across the x-axis. The direction of treatment effect is indicated by colour. The size of the p-value is indicated by the colour shade/saturation. Total number of events indicated by the circle area. Allows incorporation of labels for all events.

Adaptions considered: Could incorporate some measure of precision e.g. add 95% CI bars or shade/saturation could reflect standard error instead of the size of the p-value.

Figure A.13: Alternative volcano 2 proposed by BMJ graphic designer (WST)

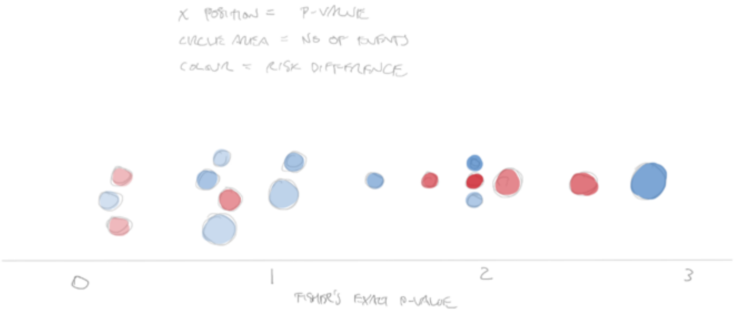


Figure description: The size of the risk difference is reflected in the colour saturation of each circle. The direction of the treatment effect is indicated by the colour of the circles. The size of the p-value is displayed across the x-axis. The total number of events is indicated by the circle area. The y-axis is not used to display a metric but instead used to stack circles/bubbles to prevent overlap.

Adaptions considered: Recommend including a legend/key to indicate which colour represents which treatment arm and size of effect that the colour shades/saturation corresponds to. Not clear how would incorporate labels.

Supplement 3 - Visualisations considered but not recommended

Figure A.14: Alternative volcano 3 proposed by BMJ graphic designer (WST)

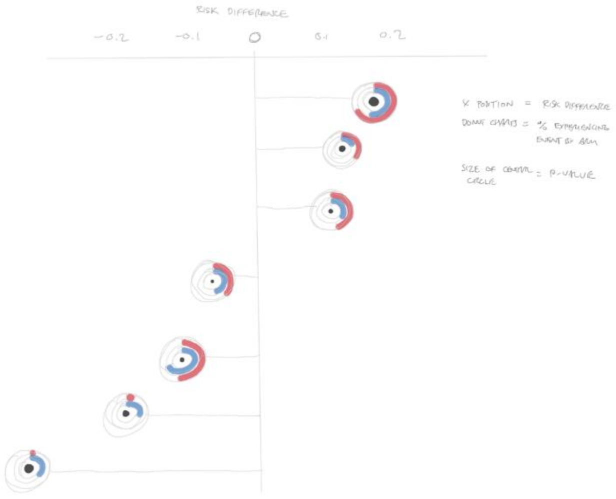

RISK DIFFERENCE
-0.2
-0.1
0
0.1
0.2
X POSITION = RISK DIFFERENCE
DONUT CHARTS = % EXPERIENCING EVENT BY ARM
SIZE OF CENTRAL CIRCLE = P-VALUE

Figure description: Shows size and direction of treatment effect across the x-axis. The size of the p-value is indicated by the size of the central black circle. The number of events is represented by the proportion of outer segments (or donuts) that are shaded, with different colours for each treatment arm. Each event takes up a ‘row’ along the y-axis. BMJ graphics designer (Will Stahl Timmins – WST) proposed this as an idea but suggests needs further refinement (please see comment below).

“If you wanted to actually show the number of people in each arm that experienced events rather than just the sum total that should be possible too. Here's one idea, but not a very good one. Concentric donut charts have issues - the inner ones tend to look smaller than the outer ones. But given time one might come up with a better one. I guess the point is, we need to work out what the most important information to show is, and then work from there…...”

Figure A.15: Tendril plot

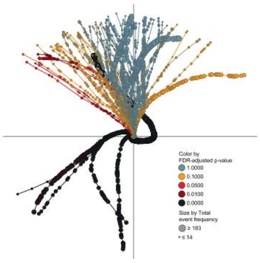


Figure 2. A tendril plot of all AEs in the RE^{2}SPOND trial. Each MedDRA PT is represented by a line (tendril) and each point is an event. Since time runs along each tendril, it is the shape that carries the important information, rather than the *x* and *y* coordinates. An event on the roflumilast treatment arm will tilt tendril direction to the left, and an event on the placebo arm will tilt tendril direction to the right. Point size is indicative of the total number of events for the type of AE in both treatment arms in the trial. The FDR adjusted Pearson's chi-squared *P*-value in each point is mapped onto a continuous color gradient.

Figure description: Each event term is represented by a line (or tendril). Each point on the line indicates the occurrence of an event. The distance from the origin indicates the time the event occurred. The direction or tilt of the line is used to indicate the treatment arm the event occurred in i.e. the line takes a unit tilt to the left for an event in the intervention arm and a unit tilt to the right for an event in the control arm. The colour of the points along the lines indicate the size of the p-value. Reprinted from *Karpefors, M. and J. Weatherall (2018). "The Tendril Plot—a novel visual summary of the incidence, significance and temporal aspects of adverse events in clinical trials." Journal of the American Medical Informatics Association* **25***(8): 1069-1073 with permission of Oxford University Press.*

Adaptions considered: This plot was initially proposed interactively so when you hover over a line it shows the event name. If it is to be used as a static plot event labels would need to be incorporated.

Supplement 3 - Visualisations considered but not recommended

Figure A.16: Heat map

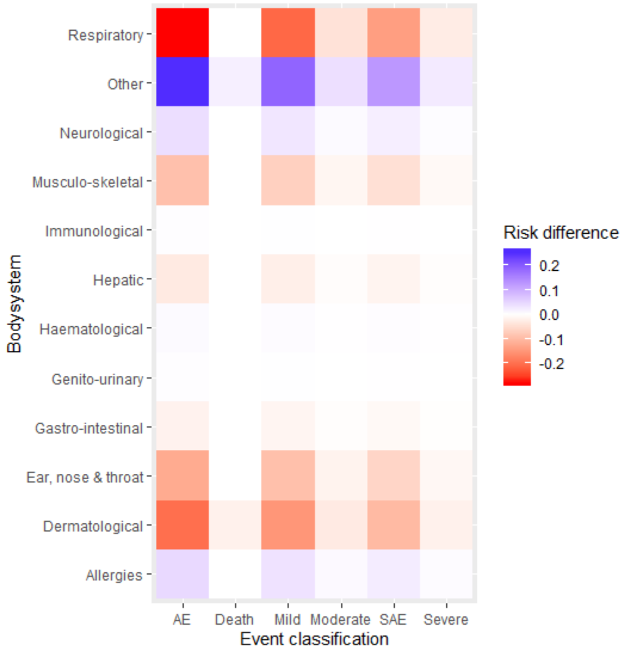


Figure description: Individual event names are displayed along the y-axis. Different event classifications such as severity ratings are displayed across the x-axis. Each x/y-axis grid position/square is coloured to represent a treatment effect for a unique event and classification. The colour of the squares/grid is used to indicate the direction of the (standardised) treatment effect. Colour saturation of the squares/grid is used to indicate size of effect for each AE. *Original plot first proposed in: Zink, R. C., et al. (2018). "Sources of Safety Data and Statistical Strategies for Design and Analysis: Clinical Trials." Therapeutic Innovation & Regulatory Science **52**(2): 141-158.*

Adaptions considered: The number of participants with each event/classification could be added to the squares/grid as per the level plot. The proportion of the square/grid coloured for each event/category combination could be used to indicate number of events as per the level plot (displayed on page 14)

Supplement 3 - Visualisations considered but not recommended

Figure A.17a: Level plot - Originally proposed for categories of abnormal blood tests

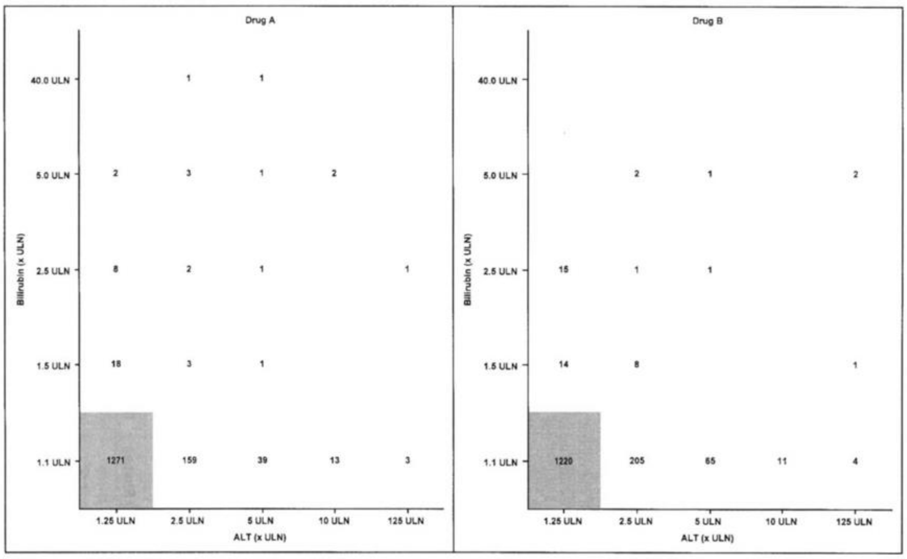


FIGURE 7. A simultaneous grade change in ALT and bilirubin.

Figure description: Displays categories of two different blood tests on the x and y-axes. Displays counts of participants in the intersection of categories. *Reprinted from: Chuang-Stein, C., et al. (2001). "Recent Advancements in the Analysis and Presentation of Safety Data." Drug Information Journal 35(2): 377-397 under the terms of the Creative Commons CC BY License.*

Adaptions considered: An adaption similar to this from Ballarini et al. could provide potentially useful modifications for the heat map. *Reprinted from: Ballarini, NM, Chiu, Y-D, König, F, Posch, M, Jaki, T. A critical review of graphics for subgroup analyses in clinical trials. Pharmaceutical Statistics. 2020; 1– 20. https://doi.org/10.1002/pst.2012 under the terms of the Creative Commons CC BY License.*

Figure A.17b Level plot – taken from Ballarini et al. for consideration of potentially useful modifications for the heat map

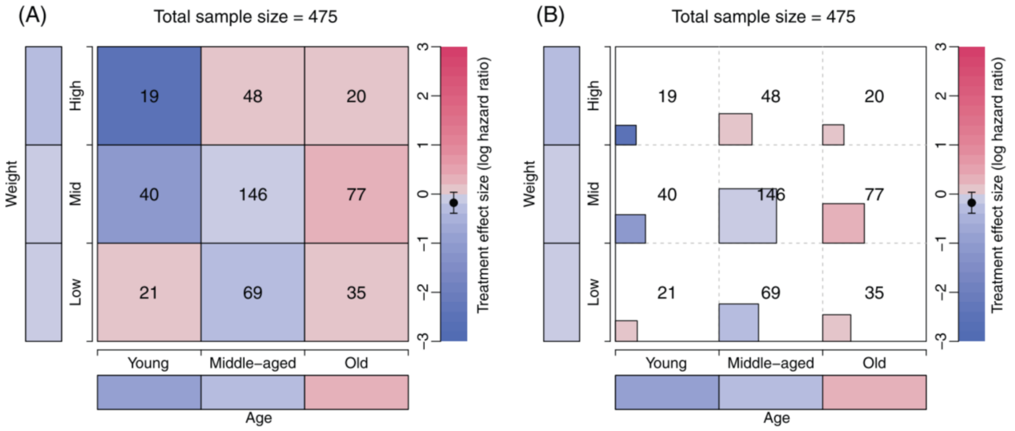


FIGURE A2 Level plots of treatment effect in terms of the log-hazard ratio across mutually disjoint subgroups defined by *age* and *weight* categorised in three levels. The cells on the bottom and the left margins correspond to the marginal subgroups defined by the levels of *age* and *weight*. In B, the area of each square inside the cells is proportional to the sample sizes, which are also displayed in the middle of the cells

Reprinted from: Ballarini, NM, Chiu, Y-D, König, F, Posch, M, Jaki, T. A critical review of graphics for subgroup analyses in clinical trials. Pharmaceutical Statistics. 2020; 1– 20, https://doi.org/10.1002/pst.2012 under the terms of the Creative Commons CC BY License.

Supplement 3 - Visualisations considered but not recommended

The following two figures have not been specifically proposed for the analysis of harm outcomes but were suggested as potentially useful plots by academic colleagues:

Figure A.18: Star plot

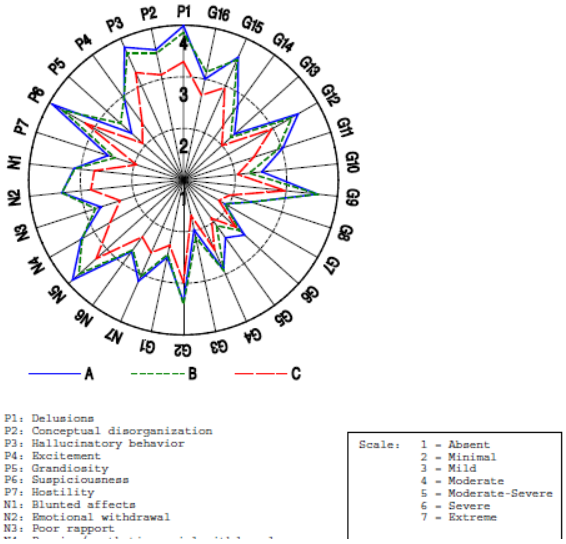


Figure description: Displays mean values for multiple clinical results (in this example each of the 30 PANSS items rated on a Likert scale). The coloured lines represent different treatment arms. Concentric reference lines included to help read off values. Could be a**dapted to** present mean grade for each AE by treatment arm. *Thanks to Steven Julious at Sheffield University for flagging this plot. Reprinted from: Squassante et al. Simple graphical methods of displaying multiple clinical results. Pharmaceut. Statist. 2006; 5: 51–60 with permission from John Wiley & Son*

Adaptions considered: Could present mean grade for each event by treatment arm

Figure A.19: Alluvial plot

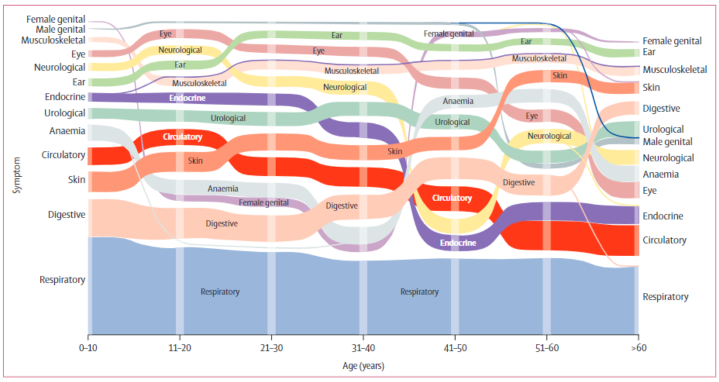


***Figure 4:* Alluvial graph of disease or symptom prevalence across decades of life**
Width of bands corresponds to relative proportion of symptoms or medical conditions, with the y axis organised to have the most prevalent conditions closest to the x axis. Associations between symptoms or conditions are represented as offshoots that connect systems. The light shaded vertical bars correspond to the decade-wise groupings used in analysis..

Figure description: Shows how the proportions experiencing an event (for multiple events) change over time. Could be adapted to show changes in severity categories over time for a single event, would be tricky to do this for multiple events. Would need to produce a separate plot for each arm to make a comparison between treatment arms. *Thanks to Marianna Nodale at Cambridge University Hospital for suggesting this image. Reprinted from: Salvi S, Apte K, Madas S, et al. Symptoms and medical conditions in 204 912 patients visiting primary health-care practitioners in India: a 1-day point prevalence study (the POSEIDON study). Lancet Glob Health. 2015;3(12):e776-e784. doi:10.1016/S2214-109X(15)00152-7 under the terms of the Creative Commons CC BY NC ND License*

Adaptions considered: Could be used to show changes in severity categories over time for a single event, would be tricky to do this for multiple events. Would need to produce a separate plot for each treatment group to make a comparison between treatments.

Supplement 3 - Visualisations considered but not recommended

Figure A.20: Histogram of counts over time

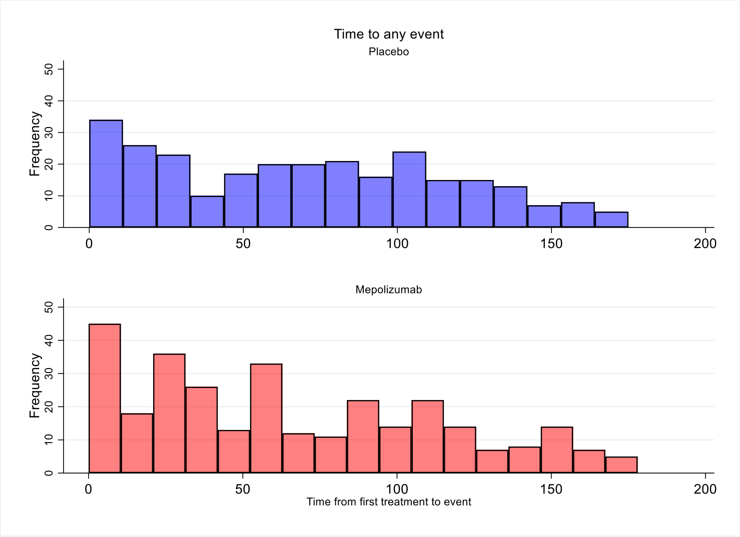


Figure description: Histogram of time of events (rather than categorising into arbitrary time periods). Not just looking at time-to-first event or maximum event, includes time of every event.

Figure A.21: Nelson-Aalen cumulative hazards

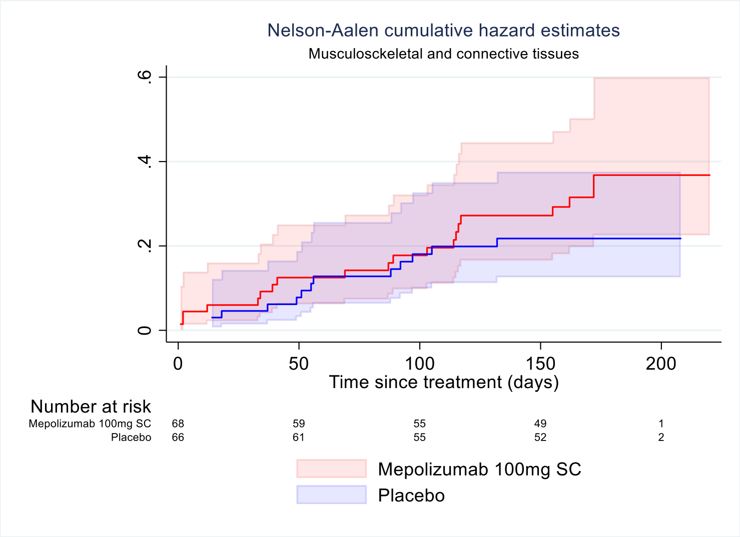


Figure description: Cumulative hazards by treatment arm with 95% confidence intervals and a table of numbers at risk

Adaptions considered: Without 95% CI &/or risk table.

Supplement 3 - Visualisations considered but not recommended

Figure A.22: Mean cumulative duration

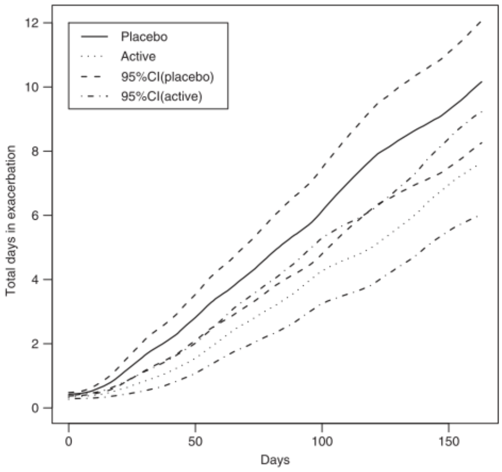


Figure 1 Mean duration of exacerbation with 95% CI based on the robust variance estimate for recurrent pulmonary exacerbations in patients with fibrosis treated with placebo and rhDNase (Fuchs et al., 1994).

Figure description: Displays the mean cumulative duration (MCD) as a function of time by treatment arm. The MCD is a non-parametric estimate of the mean cumulative duration of events per participant. Accounts for repeated occurrence of an event in a participant. Includes 95% confidence interval bands across follow-up. *Reprinted from: Wang, J. and G. Quartey (2012). "Nonparametric estimation for cumulative duration of adverse events." Biometrical Journal **54**(1): 61-74 with permission from John Wiley & Sons.*

Adaptions considered: Different colours for each treatment group

Figure A.23: Bar chart of median time to event

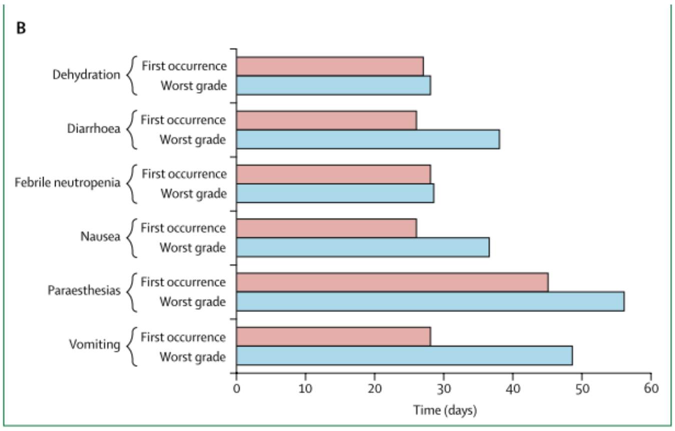


***Figure 3*: Time-to-event analyses for adverse events using the Toxicity over Time package**
(A) Time to grade 2 or worse diarrhoea in patients given FOLFOX and IROX in NCCTG N9741.[36] (B) Median time to first occurrence and worst grade toxic effects in patients given IROX in NCCTG N9741.[36] The figures capture the time profile of adverse events from these regimens. IROX=irinotecan and oxaliplatin. FOLFOX=leucovorin, fluorouracil, and oxaliplatin.

Figure description: Horizontal bar graph of median time to first (and worst grade) event. Height/length of each bar represents the median time to event. Different events are displayed along the y-axis. Time is displayed on the x-axis. Use separate bars for each treatment arm instead of first and worst event. **Caution:** This is taken from a publication in the Lancet Oncology but we think it could be very misleading since: it doesn't account for censoring or show how the denominator changes over time; and it doesn't include any information on the number of participants that have these events. *Reprinted from: Thanarajasingam G, Atherton PJ, Novotny PJ, Loprinzi CL, Sloan JA, Grothey A. Longitudinal adverse event assessment in oncology clinical trials: the Toxicity over Time (ToxT) analysis of Alliance trials NCCTG N9741 and 979254. Lancet Oncol. 2016;17(5):663-670. doi:10.1016/S1470-2045(16)00038-3 with permission from Elsevier.*

Adaptions considered: Include separate bars for each treatment group instead of first and worst event

Supplement 3 - Visualisations considered but not recommended

Figure A.24: Empirical distribution of maximum change

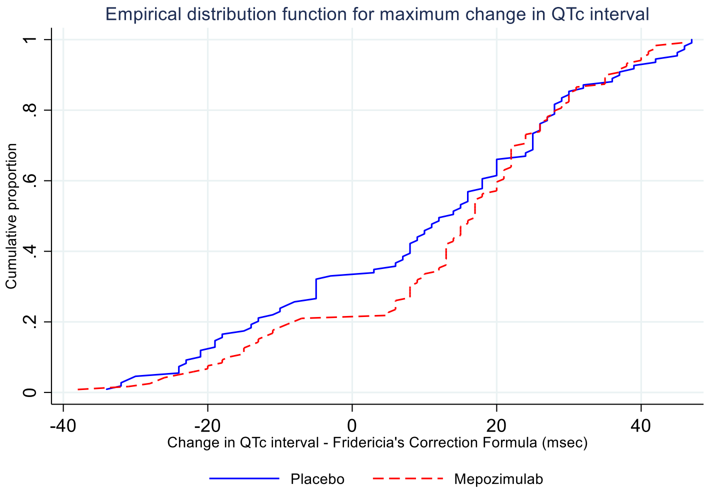


Figure description: Displays the cumulative proportion of participants on the y-axis with a change in QTc less than or equal to the corresponding value on the x-axis. Displays maximum change for each participant. Treatment arms displayed in different colours. *Original plot first proposed in: Amit, O., et al. (2008). "Graphical approaches to the analysis of safety data from clinical trials."* <u>*Pharmaceutical Statistics*</u> ***7****(1): 20-35.*

Figure A.25: Box plot of change values

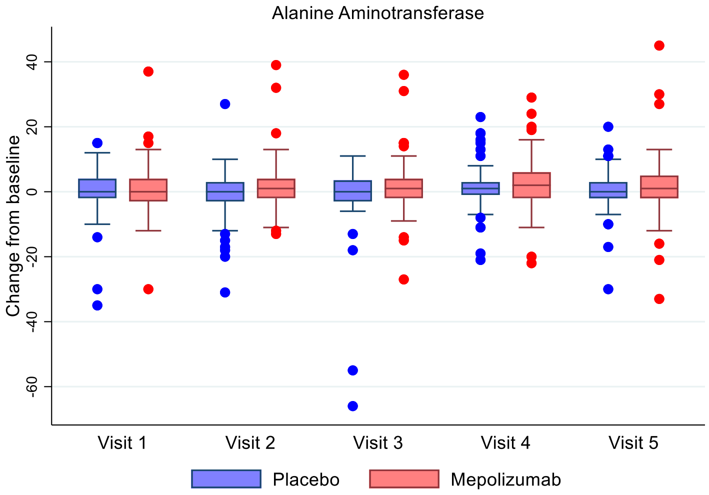


Figure description: Box plot of change from baseline across visits by treatment arm. Treatment arms displayed in different colours.

Supplement 3 - Visualisations considered but not recommended

Figure A.26: Delta plot

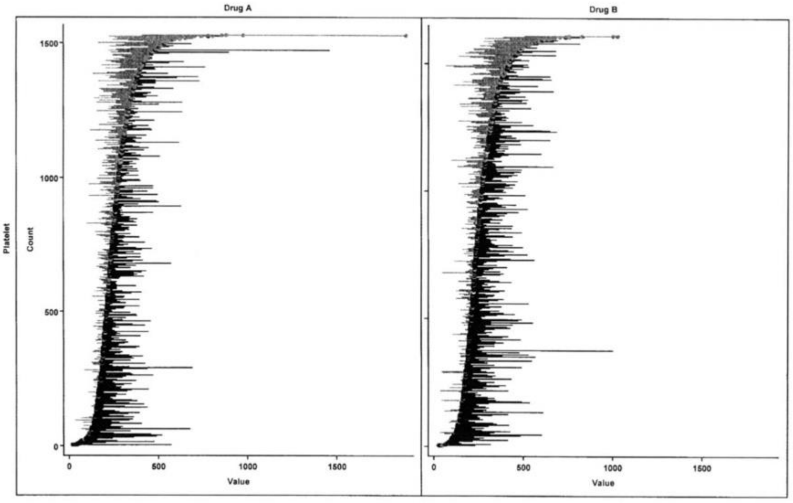
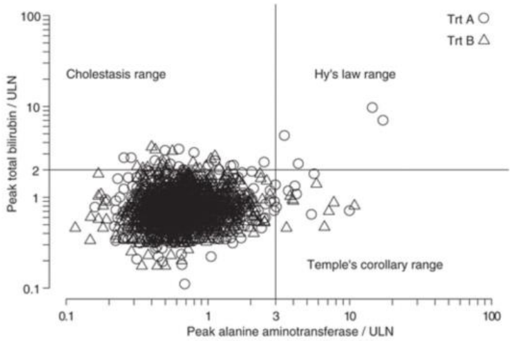


FIGURE 2a. Side-by-side Delta plots of baseline and last follow-up platelet values.

Figure description: Displays individual participant changes. The ends of each line indicate baseline and last follow-up values read from the x-axis for individual participants. Arranged according to baseline values. Y-axis tracks cumulative number of lines/participants. **Caution**: We do not find this plot very intuitive/helpful but included for comprehension. *Reprinted from: Chuang-Stein, C., et al. (2001). "Recent Advancements in the Analysis and Presentation of Safety Data."* Drug Information Journal *35(2): 377-397 under the terms of the Creative Commons CC BY License.*

Figure A.27: E-dish plot



ULN = upper limit of normal of the reference range

Figure 2 e-DISH-like plot. Notes: Plot of peak bilirubin (/ULN) *vs* peak ALT (/ULN). This figure shows peak values for total bilirubin and aminotransferases by treatment groups. Significant elevations of aminotransferases ('Temple's Corollary range') and, especially, abnormalities in the 'Hy's Law Range' should be carefully analyzed as potential signals for drug-induced liver injury [11]. Please see Refs [12] and [13] for further information on the use of this plot

Figure description: Specific scatterplot for maximum ALT, AST & Bilirubin values. Plots peak bilirubin vs peak ALT or AST. **Note**: Again, we need to consider where, if anywhere, we would advise using such an image. Perhaps better suited to monitoring of ongoing trials. *Reprinted from: Xia HA, Crowe BJ, Schriver RC, Oster M, Hall DB. Planning and core analyses for periodic aggregate safety data reviews. Clin Trials. 2011;8(2):175-182. doi:10.1177/1740774510395635 with permission from Sage Publishing*

Supplement 3 - Visualisations considered but not recommended

Figure A.28: Vector plot

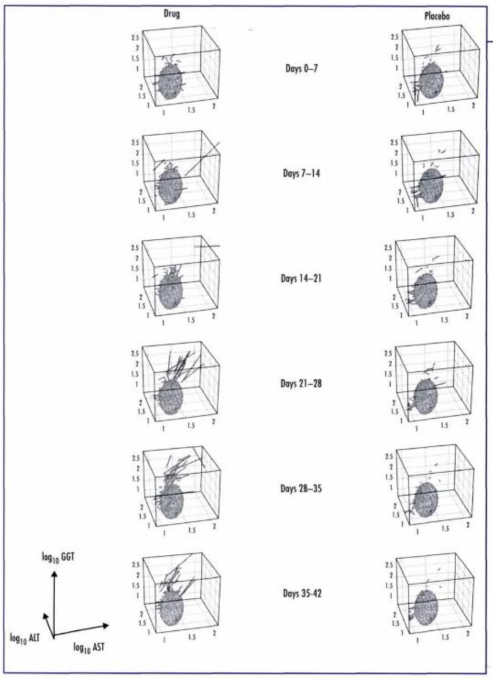


FIGURE 2

Illustration of visual vector analysis applied to ALT, AST, and GGT values simultaneously for each subject treated with placebo or active drug. In this case, the vectors for each subject move in three dimensions. The ellipsoidal surface is the 95% reference region for normal subjects. Vector movement was evident in the active-treatment group during the 14–21-day interval and becomes grossly evident in the 21–28-day interval.

Figure description: Simultaneously displays individual participant changes across three laboratory values. Grey circle indicates the 95% reference range of values for ‘normal’ subjects. **Caution**: 3D images may hide some information when viewed in a static format so we do not explore this image any further. *Reprinted from: Trost, D. C. and J. W. Freston (2008). "Vector Analysis to Detect Hepatotoxicity Signals in Drug Development." Therapeutic Innovation & Regulatory Science 42(1): 27-34 under the terms of the Creative Commons CC BY License.*

Supplement 4: Example R code and dataset to create the dot plot

Example R code

```
#########################################################################################
# R Code for Dot plot to visualize AE and harm profiles in two-arm randomised controlled trials
# Graham Wheeler and Rachel Phillips, Imperial Clinical Trials Unit, Imperial College London
# Based on code developed by Riaz Qureshi, Johns Hopkins University available https://github.com/rquresh/HarmsVisualization/blob/main/DotPlot.Rmd
#########################################################################################


### Set Working directory to be whichever folder contains the data
setwd("")
### Import the comma separate value (.csv) file
SummaryDataset <- read.csv("example_dataset_dotplot_summary_level.csv", sep=",", fileEncoding = 'UTF-8-BOM')
RDSortedBS <- SummaryDataset[order(-SummaryDataset$relrisk ),]


### Create subset of full data with the elements that we need for each half of the plot
BSRisk <- subset(RDSortedBS, select=c(event,Intervention,Placebo,relrisk))
BSRisk$event <- factor(BSRisk$event, levels = BSRisk$event[order(-BSRisk$relrisk)])
BSRiskRatio<- subset(RDSortedBS, select=c(event,relrisk,lowerCIRR,upperCIRR))
BSRiskRatio$event <- factor(BSRiskRatio$event, levels = BSRiskRatio$event[order(-BSRiskRatio$relrisk)])


### If any of the packages below have not yet been installed, use "install.packages("<package>") as below:
# install.packages("reshape")
# install.packages("ggplot2")
# install.packages("scales")


library(reshape)
ByGroup <- melt(BSRisk, id=c("event"))
ByGroup <- ByGroup[ByGroup$variable != "relrisk",]


library(ggplot2)


### "left" is a ggplot object of the group-specific risks (percentage of participants experiencing each type of event)
left <- ggplot(ByGroup, aes(x=value, y=event, fill=variable)) +
  geom_dotplot(binaxis='y', stackdir='center', dotsize = 0.5) +
  scale_fill_manual(values=c("red", "blue")) +
  ggtitle("") +
  ylab("Body system") +
  scale_x_continuous(name = "Percentage of participants (%)",  ) +
  scale_y_discrete(limits = rev(levels(ByGroup$event))) +
  theme(legend.position="bottom",
       legend.title = element_blank(),
       panel.background = element_blank(),
       panel.border = element_blank(),
       panel.grid.major.y = element_line(color = "grey", size = 0.1, linetype = 1),
       axis.ticks.x = element_line(colour = "black"),
       axis.line = element_line(color = 'black'))

### You can plot "left" on its own to check its appearance:
# left

### "right" is a ggplot object of the effect estimate and corresponding confidence interval
library(scales)
right <-  ggplot(BSRiskRatio, aes(y=event, x=relrisk, xmin=lowerCIRR, xmax=upperCIRR, fill = "Relative risk with
95% CI")) +
  ggstance::geom_pointrangeh(aes(xmin = lowerCIRR, xmax = upperCIRR)) +
  ggtitle("")+
  geom_vline(xintercept = 1, linetype = 2, colour = "blue", size = 0.75) +
  scale_x_continuous(name = "Relative risk with 95% CI",
                trans = log2_trans(),
                breaks = c(0.10,0.5,1,2, 5, 10, 50,100,220),
                labels = as.character(c(0.10,0.5,1,2, 5, 10, 50,100,220))) +
  scale_y_discrete(limits = rev(levels(BSRiskRatio$event))) +
  theme(legend.position="bottom",
       legend.title = element_blank(),
       axis.title.y = element_blank(),
       axis.text.y = element_blank(),
       panel.background = element_blank(),
       panel.border = element_blank(),
       panel.grid.major.y = element_line(color = "grey", size = 0.1, linetype = 1),
       axis.ticks.y = element_blank(),
       axis.ticks.x = element_line(colour = "black"),
        axis.text.x = element_text(angle=0, hjust=0.5, size=10),
        axis.line.x = element_line(size = 0.5, linetype = "solid", colour = "black"))


### You can plot "right" on its own to check its appearance:
# right


### This code creates the table format for the number of patients experiencing an event per arm,
### or the total number of events of each type per arm


tab_base <- ggplot(RDSortedBS, aes(y=event)) +
  ylab(NULL) + xlab(" ") +   scale_y_discrete(limits = rev(levels(BSRiskRatio$event))) +
  theme(plot.title = element_text(hjust = 0.5, size=12), ## centering title on text
        axis.text.x=element_text(color="white"), ## need text to be printed so it stays aligned with figure but white so it's invisible
        axis.line=element_blank(),
        axis.text.y=element_blank(),axis.ticks=element_blank(),
        axis.title.y=element_blank(),legend.position="bottom", legend.title = element_blank(),
        panel.background=element_blank(),panel.border=element_blank(),panel.grid.major=element_blank(),
        panel.grid.minor=element_blank(),plot.background=element_blank())


### Tables of number of participants experiencing each harm type for Intervention (I) and Placebo (P) arms
t_I_n<-tab_base + geom_text(aes(x=1, label = eventn1, hjust = "middle")) + ggtitle(expression("I"["n"]))
t_P_n<-tab_base + geom_text(aes(x=1, label = eventn2, hjust = "middle")) + ggtitle(expression("P"["n"]))


### Tables of total number of events per harm type for Intervention (I) and Placebo (P) arms
t_I_event<-tab_base + geom_text(aes(x=1, label = n_events1, hjust = "middle")) + ggtitle(expression("I"["event"]))
t_P_event<-tab_base + geom_text(aes(x=1, label = n_events2, hjust = "middle")) + ggtitle(expression("P"["event"]))



### Now put "left", "right", and tables of your choice together in one figure.
### Again, install any of the packages below if they have not yet been installed:
# install.packages("ggpubr")
# install.packages("cowplot")
library(ggpubr)
library(cowplot)
```

Supplement 4: Example R code and dataset to create the dot plot

```
### Dot plot Version 1 - only number of participants with each events

DotPlot <- plot_grid(left, right, t_I_n, t_P_n, nrow = 1, align = "h", rel_widths = c(3,2,1,1), axis = "b")

### NB: may take a few seconds to generate

annotate_figure(DotPlot,

        bottom = text_grob(bquote("I:Intervention (N = "*.(RDSortedBS[1,"N1"])*"), P:Placebo (N = "*.(RDSortedBS[1,"N2"])*"); X"["n"]*"= number of participants in arm X with AE"),

                color = "black", face = "bold", size = 10),

        top = text_grob("", color = "black", face = "bold", size = 10))


### Dot plot Version 2 - number of participants with each event and total number of events

DotPlot <- plot_grid(left, right, t_I_n, t_I_event, t_P_n, t_P_event, nrow = 1, align = "h", rel_widths = c(3,2,0.5,0.5,0.5,0.5), axis = "b")

### NB: may take a few seconds to generate

annotate_figure(DotPlot,

        bottom = text_grob(bquote("I:Intervention (N = "*.(RDSortedBS[1,"N1"])*"), P:Placebo (N = "*.(RDSortedBS[1,"N2"])*"); X"["n"]*"= number of participants in arm X with AE, X"["event"]*"= number of AEs in arm X"),

                color = "black", face = "bold", size = 10),

        top = text_grob("", color = "black", face = "bold", size = 10))


#######

# END #

#######
```

Supplement 4: Example R code and dataset to create the dot plot

Example dataset

r1	eventn1	N1	r2	eventn2	N2	Intervention	Placebo	risk_diff	seRD	lowerRD	upperRD	n_events	n_events1	n_events2	p_val	log_p_val	event	relrisk	logRR	stderrRR	loglowerCIRR	logupperCIRR	lowerCIRR	upperCIRR
0.766667	23	30	1	31	31	76.66666	100	-0.23333	0.07722	-0.38468	-0.08198	158	68	90	0.004666	2.331021	Blood and lymphatic	0.766667	-0.2657	0.100722	-0.46312	-0.06829	0.629318	0.933991
0.233333	7	30	0.258065	8	31	23.33333	25.80645	-0.02473	0.110179	-0.24068	0.191219	23	14	9	1	2.18E-14	Dermatological	0.904167	-0.10074	0.44974	-0.98223	0.780748	0.374475	2.183105
0.1	3	30	0.096774	3	31	10	9.67742	0.003226	0.076287	-0.1463	0.152748	6	3	3	1	-4.82E-16	Eyes, ear, nose, thr	1.033333	0.03279	0.77529	-1.48678	1.552359	0.2261	4.722598
0.533333	16	30	0.83871	26	31	53.33333	83.87096	-0.30538	0.112517	-0.52591	-0.08484	94	29	65	0.013428	1.871977	Gastrointestinal	0.635897	-0.45272	0.18807	-0.82133	-0.0841	0.439844	0.919338
0.166667	5	30	0.032258	1	31	16.66667	3.225807	0.134409	0.075078	-0.01274	0.281561	8	7	1	0.103516	0.984994	Infection	5.166667	1.642228	1.065086	-0.44534	3.729797	0.640606	41.67064
0.233333	7	30	0.193548	6	31	23.33333	19.35484	0.039785	0.104872	-0.16576	0.245333	29	14	15	0.762211	0.117925	Neurological	1.205556	0.186941	0.493895	-0.78109	1.154975	0.457905	3.173944
0.4	12	30	0.225806	7	31	40	22.58065	0.174194	0.116787	-0.05471	0.403097	23	13	10	0.173665	0.760288	Other	1.771429	0.571786	0.400748	-0.21368	1.357253	0.807607	3.885504
0.133333	4	30	0.096774	3	31	13.33333	9.67742	0.036559	0.081679	-0.12353	0.19665	14	10	4	0.707182	0.150469	Psychiatric	1.377778	0.320472	0.719543	-1.08983	1.730776	0.336273	5.645032
0.033333	1	30	0.129032	4	31	3.333333	12.90323	-0.0957	0.068552	-0.23006	0.038662	9	3	6	0.353987	0.451012	Renal and urinary	0.258333	-1.3535	1.088305	-3.48658	0.779574	0.030605	2.180543
0.033333	1	30	0.129032	4	31	3.333333	12.90323	-0.0957	0.068552	-0.23006	0.038662	6	1	5	0.353987	0.451012	Respiratory	0.258333	-1.3535	1.088305	-3.48658	0.779574	0.030605	2.180543

Supplement 5: Table and figures summarising initial appraisals of all plots by outcome type

Table A.2a: Plots suitable for **Multiple Binary Outcomes** – summary of scores

Appraisal criteria	**Volcano**		**Alternative volcano 1**		**Alternative volcano 2**		**Alternative volcano 3**		**Dot plot**		**Bar**	
	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)
		Median (Min, Max)		Median (Min, Max)		Median (Min, Max)		Median (Min, Max)		Median (Min, Max)		Median (Min, Max)
1.Effect size	21	4.0 (1.0)	21	4.0 (0.8)	21	2.3 (1.2)	15	3.2 (1.3)	21	4.3 (1.1)	21	2.8 (1.1)
		4 (1, 5)		4 (3, 5)		2 (1, 5)		4 (1, 5)		5 (1, 5)		3 (1, 5)
2.Direction of effect	21	4.1 (1.1)	21	4.0 (1.2)	20	2.9 (1.4)	14	2.8 (1.7)	21	4.5 (0.8)	21	4.0 (1.1)
		4 (1, 5)		4 (1, 5)		3 (1, 5)		3 (1, 5)		5 (2, 5)		4 (1, 5)
3.Uncertainty	21	1.8 (1.1)	21	1.8 (0.7)	21	2.0 (1.0)	14	1.3 (0.5)	21	4.4 (1.0)	21	1.3 (0.6)
		1 (1, 5)		2 (1, 3)		2 (1, 5)		1 (1, 2)		5 (1, 5)		1 (1, 3)
4.Supplementary data needed	21	2.1 (1.1)	21	2.5 (1.1)	21	2.0 (1.1)	13	2.2 (1.1)	21	3.8 (1.2)	21	2.6 (1.1)
		2 (1, 4)		2 (1, 5)		2 (1, 5)		2 (1, 4)		4 (1, 5)		2 (1, 5)
5.Understandable	21	2.9 (0.8)	21	3.0 (1.1)	21	2.3 (1.0)	14	1.5 (0.7)	21	4.2 (0.8)	21	4.8 (0.4)
		3 (2, 4)		3 (1, 5)		2 (1, 4)		1 (1, 3)		4 (2, 5)		5 (4, 5)
6.Understandable non-stats	21	2.3 (0.9)	21	2.9 (1.0)	21	2.3 (0.9)	14	1.4 (0.6)	21	4.0 (0.9)	20	4.7 (0.6)
		2 (1, 4)		3 (1, 4)		2 (1, 4)		1 (1, 3)		4 (2, 5)		5 (3, 5)
7.Multi-arm studies	21	2.0 (0.7)	21	2.1 (0.9)	21	3.8 (1.0)	14	1.9 (1.0)	21	3.0 (1.0)	21	4.7 (0.7)
		2 (1, 3)		2 (1, 4)		4 (1, 5)		2 (1, 4)		3 (1, 5)		5 (2, 5)
8.Limits numbers	21	3.2 (0.8)	21	3.9 (0.8)	21	3.7 (1.1)	14	3.1 (1.2)	21	3.7 (0.8)	21	3.9 (0.7)
		3 (2, 5)		4 (2, 5)		4 (1, 5)		3 (1, 5)		4 (2, 5)		4 (3, 5)
9.Overall score*	**21**	**19.2 (4.5)**	**21**	**20.3 (4.0)**	**21**	**17.3 (5.0)**	**21**	**9.6 (7.6)**	**21**	**28.1 (5.0)**	**21**	**24.6 (3.2)**
		19 (11, 27)		**20 (12, 28)**		**17 (7, 28)**		**11 (0, 23)**		**29 (15, 34)**		**26 (18, 30)**
10. Suitable for publication	21	3.3 (1.1)	21	3.6 (1.0)	19	2.5 (1.4)	14	1.7 (1.1)	20	4.3 (0.9)	20	3.8 (1.0)
		3 (1, 5)		4 (1, 5)		2 (1, 5)		1 (1, 4)		5 (2, 5)		4 (2, 5)
11. Suitable for final report	20	3.5 (1.1)	20	3.7 (1.0)	19	2.5 (1.4)	14	1.8 (1.3)	20	4.3 (0.7)	20	4.0 (0.9)
		4 (1, 5)		4 (1, 5)		2 (1, 5)		1 (1, 5)		5 (3, 5)		4 (2, 5)
12. Suitable for interim analysis	20	3.2 (1.2)	20	3.5 (1.1)	19	2.4 (1.3)	14	1.5 (0.9)	20	4.3 (0.7)	20	4.1 (0.7)
		3 (1, 5)		4 (1, 5)		2 (1, 5)		1 (1, 4)		4 (3, 5)		4 (3, 5)
13.Exploratory analysis	20	3.7 (0.6)	20	3.4 (0.8)	19	2.8 (0.9)	14	1.9 (1.1)	19	3.8 (0.8)	19	3.7 (1.1)
		4 (3, 5)		4 (2, 5)		3 (1, 4)		2 (1, 4)		4 (2, 5)		4 (1, 5)
14.Explanatory analysis	20	3.1 (0.7)	20	3.3 (0.9)	19	2.5 (1.0)	14	1.9 (1.1)	19	4.1 (0.8)	19	3.5 (0.9)
		3 (2, 4)		3 (2, 5)		3 (1, 4)		2 (1, 4)		4 (3, 5)		3 (2, 5)
Ranking	**18**	**5.6 (2.1)**	**18**	**4.8 (1.8)**	**18**	**6.6 (2.9)**	**14**	**9.8 (2.3)**	**20**	**1.6 (1.6)**	**17**	**3.8 (1.9)**
		5 (2, 10)		**5 (2, 9)**		**7 (1, 12)**		**11 (4, 12)**		**1 (1, 7)**		**4 (1, 9)**

* Overall score is the sum total of questions 1-7

Supplement 5: Table and figures summarising initial appraisals of all plots by outcome type

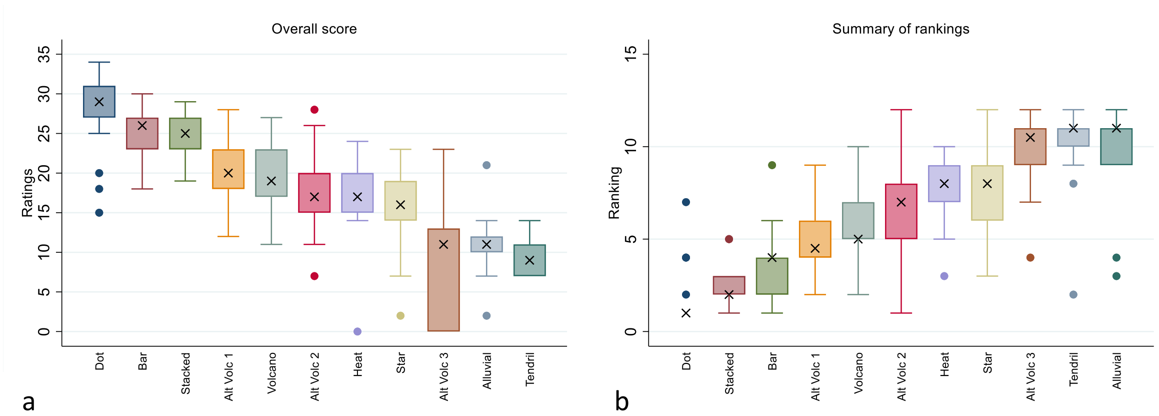
Table A.2b: Plots suitable for **Multiple Binary Outcomes** – summary of scores

Question	**Tendril**		**Heat map**		**Stacked bar chart**		**Stacked bar chart - counts**		**Star**		**Alluvial**	
	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)
		Median (Min, Max)		Median (Min, Max)		Median (Min, Max)		Median (Min, Max)		Median (Min, Max)		Median (Min, Max)
1.Effect size	21	1.3 (0.6)	20	3.0 (0.7)	21	3.1 (1.2)	12	2.4 (1.1)	21	2.0 (0.9)	21	1.4 (0.6)
		1 (1, 3)		3 (2, 4)		3 (1, 5)		3 (1, 4)		2 (1, 4)		1 (1, 3)
2.Direction of effect	21	1.6 (1.0)	20	3.6 (1.0)	21	4.1 (0.9)	12	3.3 (1.2)	21	2.7 (1.0)	21	1.3 (0.6)
		1 (1, 5)		4 (1, 5)		4 (1, 5)		4 (1, 5)		3 (1, 4)		1 (1, 3)
3.Uncertainty	21	1.2 (0.5)	20	1.3 (0.7)	21	1.4 (0.6)	12	1.3 (0.5)	20	1.1 (0.4)	20	1.1 (0.4)
		1 (1, 3)		1 (1, 4)		1 (1, 3)		1 (1, 2)		1 (1, 2)		1 (1, 2)
4.Supplementary data needed	21	1.6 (1.2)	20	2.1 (0.8)	21	3.2 (0.9)	12	3.0 (1.0)	20	1.9 (1.1)	20	1.9 (1.4)
		1 (1, 5)		2 (1, 4)		3 (2, 5)		3 (2, 5)		2 (1, 5)		1 (1, 5)
5.Understandable	21	1.2 (0.5)	20	3.1 (1.0)	21	4.7 (0.5)	12	4.4 (0.8)	20	2.3 (0.9)	20	2.1 (1.0)
		1 (1, 3)		3 (2, 5)		5 (4, 5)		5 (3, 5)		2 (1, 4)		2 (1, 4)
6.Understandable non-stats	21	1.0 (0.2)	20	2.9 (1.0)	21	4.5 (0.6)	12	4.1 (1.0)	20	1.9 (0.7)	20	1.9 (0.9)
		1 (1, 2)		3 (2, 5)		5 (3, 5)		4 (2, 5)		2 (1, 3)		2 (1, 4)
7.Multi-arm studies	21	1.4 (0.7)	20	1.9 (0.9)	21	4.1 (1.2)	12	4.3 (1.2)	20	3.9 (1.4)	20	1.6 (0.7)
		1 (1, 3)		2 (1, 4)		5 (1, 5)		5 (2, 5)		4 (1, 5)		1 (1, 3)
8.Limits numbers	21	3.3 (1.2)	19	3.6 (1.0)	21	3.5 (0.7)	11	3.5 (0.7)	19	2.9 (1.0)	20	2.5 (1.1)
		4 (1, 5)		4 (1, 5)		3 (2, 5)		3 (3, 5)		3 (1, 5)		3 (1, 4)
9.Overall score*	**21**	**9.3 (2.4)**	**21**	**17.1 (5.1)**	**21**	**25.1 (2.8)**	**21**	**13.0 (12.0)**	**21**	**15.4 (5.0)**	**21**	**10.9 (3.5)**
		9 (7, 14)		**17 (0, 24)**		**25 (19, 29)**		**15 (0, 28)**		**16 (2, 23)**		**11 (2, 21)**
10. Suitable for publication	20	1.5 (0.9)	20	2.8 (0.9)	21	4.0 (0.9)	11	3.5 (1.1)	20	2.2 (1.2)	19	1.7 (1.1)
		1 (1, 4)		3 (1, 4)		4 (2, 5)		4 (2, 5)		2 (1, 5)		1 (1, 4)
11. Suitable for final report	20	1.8 (1.1)	20	2.7 (0.9)	21	4.0 (0.8)	11	3.6 (1.0)	20	2.4 (1.3)	19	2.0 (1.1)
		1 (1, 4)		3 (1, 4)		4 (2, 5)		4 (2, 5)		2 (1, 5)		2 (1, 4)
12. Suitable for interim analysis	20	2.1 (1.4)	20	2.5 (0.9)	21	4.1 (0.8)	11	3.6 (1.0)	20	2.3 (1.2)	19	2.1 (1.2)
		2 (1, 5)		3 (1, 4)		4 (2, 5)		4 (2, 5)		2 (1, 5)		2 (1, 4)
13.Exploratory analysis	20	3.0 (1.2)	19	3.1 (1.0)	20	4.0 (0.9)	11	3.5 (1.1)	19	2.9 (1.1)	19	2.8 (1.5)
		3 (1, 5)		3 (1, 4)		4 (2, 5)		4 (2, 5)		3 (1, 5)		3 (1, 5)
14.Explanatory analysis	19	1.7 (1.0)	19	2.6 (1.0)	20	4.0 (1.1)	11	3.3 (1.3)	19	2.2 (1.1)	18	1.7 (0.8)
		1 (1, 5)		3 (1, 5)		4 (1, 5)		3 (1, 5)		2 (1, 5)		2 (1, 3)
Ranking	**18**	**10.2 (2.3)**	**17**	**7.5 (1.8)**	**19**	**2.4 (1.1)**	**11**	**4.9 (2.3)**	**17**	**8.0 (2.3)**	**17**	**9.9 (2.6)**
		11 (2, 12)		**8 (3, 10)**		**2 (1, 5)**		**4 (2, 10)**		**8 (3, 12)**		**11 (3, 12)**

* Overall score is the sum total of questions 1-7

Supplement 5: Table and figures summarising initial appraisals of all plots by outcome type

Figure A.29: Multiple binary outcomes



a. Box plot of overall scores ordered by highest to lowest mean values (higher scores indicate better performance). **b. Box plot of rankings** ordered by best to worst mean rank (lower ranking indicates preferred plot).

Note: X indicates median values. Excludes summary for stacked bar chart of counts as only limited numbers scored this plot

Supplement 5: Table and figures summarising initial appraisals of all plots by outcome type

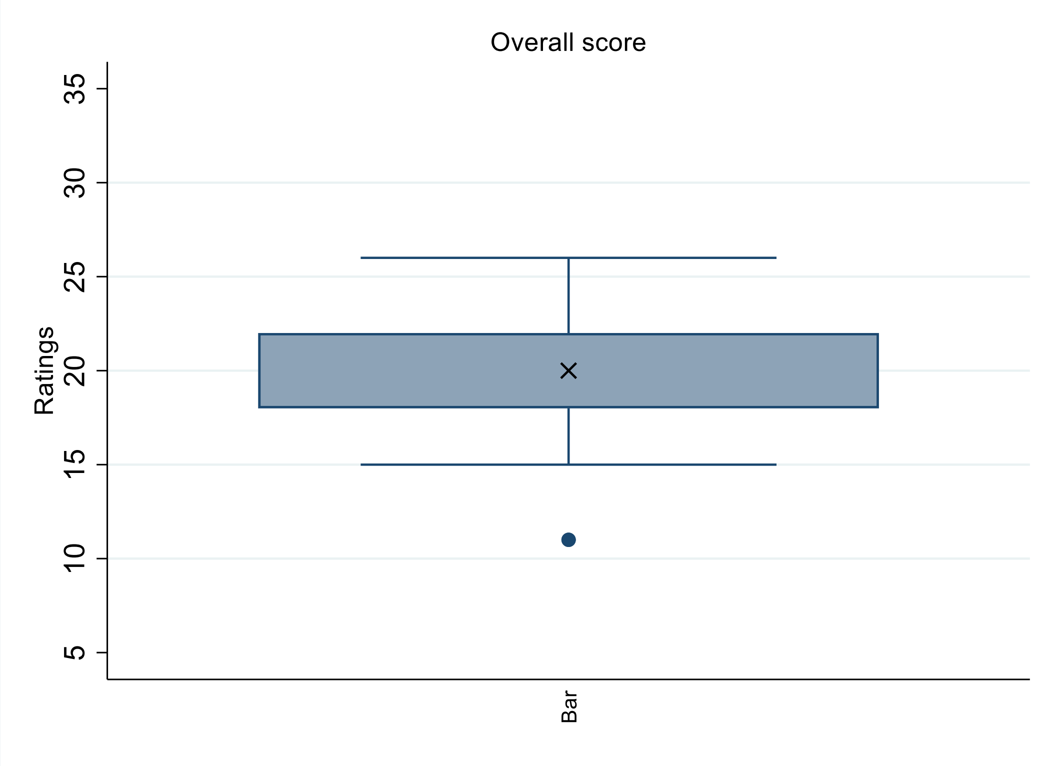
Table A.3: Plots suitable for **Single Binary Outcomes** – summary of scores

Question	**Bar chart**	
	n	Mean (SD) Median (Min, Max)
1.Effect size	23	2.1 (0.9)
		2 (1, 4)
2.Direction of effect	23	2.3 (1.0)
		2 (1, 4)
3.Uncertainty	23	1.1 (0.3)
		1 (1, 2)
4.Supplementary data needed	23	2.5 (1.2)
		2 (1, 4)
5.Understandable	23	4.0 (1.2)
		4 (1, 5)
6.Understandable non-stats	23	3.9 (0.9)
		4 (2, 5)
7.Multi-arm studies	23	3.9 (0.9)
		4 (2, 5)
8.Limits numbers	23	3.2 (1.2)
		3 (1, 5)
9.Overall score	**23**	**19.8 (3.3)**
		20 (11, 26)
10. Suitable for publication	22	3.1 (0.9)
		3 (1, 4)
11. Suitable for final report	22	3.4 (1.0)
		4 (1, 4)
12. Suitable for interim analysis	22	3.5 (1.0)
		4 (1, 5)
13.Exploratory analysis	22	3.1 (1.0)
		3 (1, 5)
14.Explanatory analysis	22	3.0 (1.0)
		3 (1, 5)
Ranking	**6**	**1.0 (0.0)**
		1 (1, 1)

* Overall score is the sum total of questions 1-7

Supplement 5: Table and figures summarising initial appraisals of all plots by outcome type

Figure A.30: Single binary outcomes



Box plot of overall scores, ordered by highest to lowest mean values (higher scores indicate better performance).

Note: X indicates median values.

Supplement 5: Table and figures summarising initial appraisals of all plots by outcome type

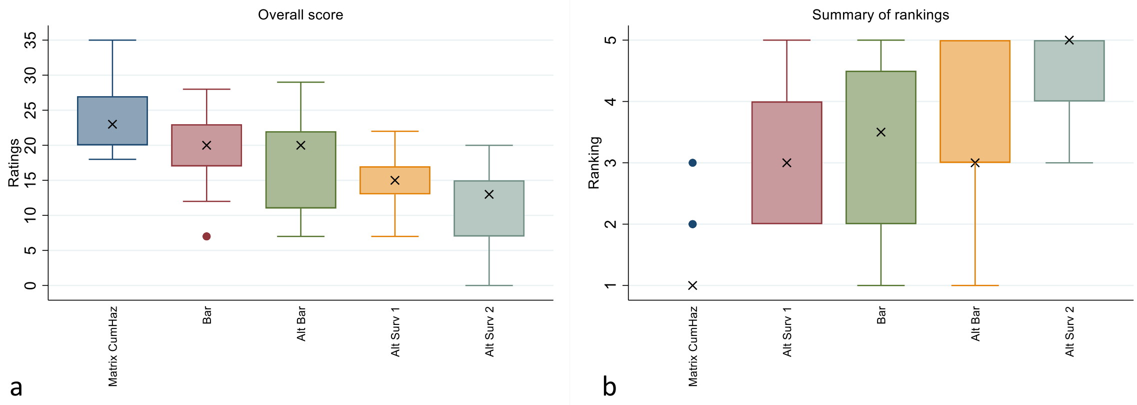
Table A.4: Plots suitable for **Multiple Time-to-Event Outcomes** – summary of scores

Question	**Matrix of cumulative hazards**		**Bar chart**		**Alternative bar chart**		**Alternative survival plot 1**		**Alternative survival plot 2**	
	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)
		Median (Min, Max)		Median (Min, Max)		Median (Min, Max)		Median (Min, Max)		Median (Min, Max)
1.Effect size	21	3.4 (1.0)	21	2.4 (1.2)	21	2.3 (1.3)	21	2.3 (0.8)	18	1.6 (0.8)
		4 (2, 5)		2 (1, 4)		2 (1, 4)		2 (1, 4)		1 (1, 4)
2.Direction of effect	21	3.9 (0.8)	21	3.0 (1.3)	21	2.8 (1.1)	21	2.9 (1.2)	18	1.4 (0.6)
		4 (2, 5)		3 (1, 5)		3 (1, 4)		3 (1, 5)		1 (1, 3)
3.Uncertainty	21	2.4 (1.5)	21	1.4 (0.8)	21	1.5 (0.8)	21	1.1 (0.3)	18	1.2 (0.4)
		2 (1, 5)		1 (1, 4)		1 (1, 3)		1 (1, 2)		1 (1, 2)
4.Supplementary data needed	20	2.5 (1.1)	21	2.1 (1.1)	21	1.9 (0.8)	21	2.0 (1.1)	18	1.7 (1.1
		2 (1, 5)		2 (1, 4)		2 (1, 3)		2 (1, 5)		1 (1, 5)
5.Understandable	21	4.1 (0.7)	21	3.2 (1.1)	21	2.7 (1.4)	21	2.2 (1.0)	18	2.5 (1.2)
		4 (3, 5)		3 (1, 5)		3 (1, 5)		2 (1, 5)		3 (1, 4)
6.Understandable non-stats	20	3.2 (1.0)	21	3.0 (1.1)	21	2.3 (1.2)	21	2.0 (1.0)	18	2.1 (1.2)
		3 (1, 5)		3 (1, 5)		2 (1, 5)		2 (1, 5)		2 (1, 4)
7.Multi-arm studies	21	4.5 (0.6)	21	3.6 (1.3)	21	3.7 (1.4)	21	1.9 (0.9)	18	2.4 (1.2)
		5 (3, 5)		4 (1, 5)		4 (1, 5)		2 (1, 4)		3 (1, 5)
8.Limits numbers	21	2.3 (1.1)	21	3.0 (1.0)	19	3.8 (1.3)	21	3.2 (1.1)	18	2.3 (1.0)
		2 (1, 5)		3 (1, 4)		4 (1, 5)		3 (1, 5)		3 (1, 4)
9.Overall score	**21**	**24.0 (4.8)**	**21**	**19.0 (5.4)**	**21**	**17.5 (6.6)**	**21**	**14.6 (4.2)**	**21**	**11.2 (6.1)**
		23 (18, 35)		**20 (7, 28)**		**20 (7, 29)**		**15 (7, 22)**		**13 (0, 20)**
10. Suitable for publication	20	3.4 (1.1)	20	2.2 (1.1)	20	2.2 (1.4)	20	1.8 (0.9)	18	1.9 (1.0)
		4 (1, 5)		2 (1, 4)		2 (1, 5)		2 (1, 4)		2 (1, 4)
11. Suitable for final report	20	3.8 (1.0)	20	2.3 (1.3)	20	2.3 (1.3)	20	1.8 (0.9)	18	1.9 (1.0)
		4 (1, 5)		2 (1, 4)		2 (1, 5)		2 (1, 4)		2 (1, 4)
12. Suitable for interim analysis	20	3.8 (1.0)	20	2.3 (1.4)	20	2.3 (1.3)	20	2.2 (1.4)	18	2.1 (1.0)
		4 (1, 5)		2 (1, 5)		2 (1, 5)		2 (1, 5)		2 (1, 4)
13.Exploratory analysis	20	3.7 (1.1)	20	2.5 (1.3)	20	2.6 (1.3)	20	3.1 (1.3)	17	2.4 (1.4)
		4 (1, 5)		3 (1, 4)		3 (1, 5)		3 (1, 5)		2 (1, 5)
14.Explanatory analysis	20	3.5 (1.1)	20	2.2 (1.3)	20	2.4 (1.2)	20	2.4 (1.0)	17	2.2 (1.1)
		4 (1, 5)		2 (1, 4)		3 (1, 4)		2 (1, 4)		2 (1, 4)
Ranking	**18**	**1.3 (0.6)**	**16**	**3.4 (1.3)**	**17**	**3.5 (1.3)**	**16**	**3.1 (1.1)**	**15**	**4.5 (0.7)**
		1 (1, 3)		**4 (1, 5)**		**3 (1, 5)**		**3 (2, 5)**		**5 (3, 5)**

* Overall score is the sum total of questions 1-7

Supplement 5: Table and figures summarising initial appraisals of all plots by outcome type

Figure A.31: Multiple time-to-event outcomes



a. Box plot of overall scores ordered by highest to lowest mean values (higher scores indicate better performance). **b. Box plot of rankings** ordered by best to worst mean rank (lower ranking indicates preferred plot).
Note: X indicates median values.

Supplement 5: Table and figures summarising initial appraisals of all plots by outcome type

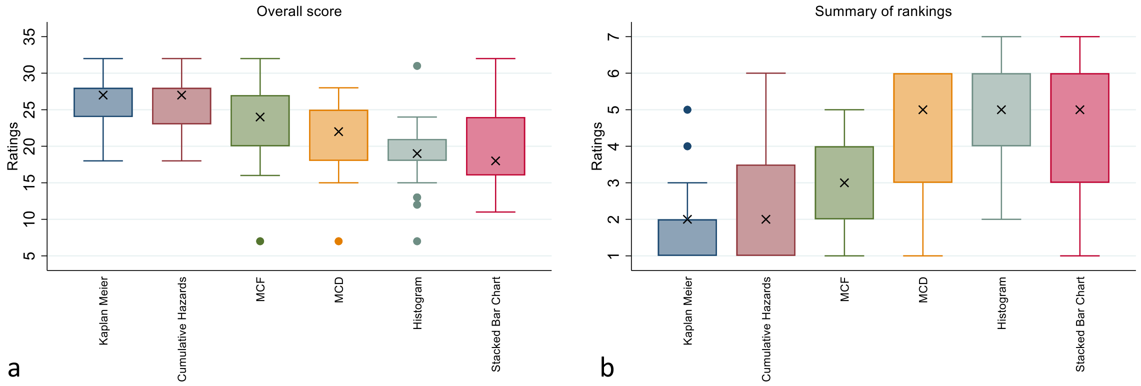
Table A.5: Plots suitable for **Single TTE Outcomes** – summary of scores

Question	**Cumulative Hazard**		**Kaplan Meier**		**Mean Cumulative Function**		**Mean Cumulative Duration**		**Stacked bar chart over time**		**Histogram of counts over time**	
	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)
		Median (Min, Max)		Median (Min, Max)		Median (Min, Max)		Median (Min, Max)		Median (Min, Max)		Median (Min, Max)
1.Effect size	23	3.1 (1.2)	23	3.2 (1.2)	23	3.0 (1.2)	23	2.6 (1.2)	23	2.2 (1.3)	23	2.0 (1.1)
		3 (1, 5)		4 (1, 5)		3 (1, 5)		3 (1, 4)		2 (1, 5)		2 (1, 5)
2.Direction of effect	23	3.8 (1.1)	23	3.8 (1.1)	23	3.7 (1.1)	23	3.2 (1.2)	23	2.7 (1.3)	23	2.5 (1.2)
		4 (1, 5)		4 (1, 5)		4 (1, 5)		3 (1, 5)		3 (1, 5)		2 (1, 4)
3.Uncertainty	23	4.0 (1.1)	23	4.0 (1.0)	23	3.6 (0.9)	23	3.7 (1.0)	23	1.2 (0.5)	23	1.1 (0.5)
		4 (1, 5)		4 (1, 5)		4 (1, 5)		4 (1, 5)		1 (1, 3)		1 (1, 3)
4.Supplementary data needed	23	3.7 (0.9)	23	3.8 (0.9)	23	2.9 (1.0)	23	2.2 (1.1)	23	2.6 (1.2)	23	2.6 (1.3)
		4 (2, 5)		4 (2, 5)		3 (1, 4)		2 (1, 4)		3 (1, 5)		3 (1, 4)
5.Understandable	23	4.2 (0.7)	23	4.3 (0.7)	23	3.2 (1.1)	23	2.8 (1.1)	23	3.4 (1.2)	23	3.7 (1.1)
		4 (3, 5)		4 (3, 5)		3 (1, 5)		3 (1, 5)		4 (1, 5)		4 (1, 5)
6.Understandable non-stats	23	3.1 (0.9)	23	3.5 (0.8)	23	2.7 (1.1)	23	2.3 (1.0)	23	3.2 (1.1)	23	3.4 (1.2)
		3 (1, 5)		4 (2, 5)		3 (1, 4)		2 (1, 4)		3 (1, 5)		3 (1, 5)
7.Multi-arm studies	21	3.9 (0.5)	21	3.9 (0.5)	21	3.8 (0.9)	21	3.5 (0.9)	21	3.9 (0.7)	21	3.4 (0.9)
		4 (3, 5)		4 (3, 5)		4 (1, 5)		4 (1, 5)		4 (3, 5)		4 (1, 5)
8.Limits numbers	23	3.1 (1.4)	23	3.1 (1.5)	23	3.4 (1.5)	23	3.2 (1.5)	22	2.8 (1.2)	23	3.5 (1.5)
		3 (1, 5)		3 (1, 5)		4 (1, 5)		3 (1, 5)		3 (1, 5)		4 (1, 5)
9.Overall score	**21**	**26.1 (3.8)**	**21**	**26.7 (3.6)**	**21**	**23.2 (5.6)**	**21**	**21.0 (5.1)**	**21**	**19.8 (5.6)**	**21**	**19.1 (5.2)**
		27 (18, 32)		**27 (18, 32)**		**24 (7, 32)**		**22 (7, 28)**		**18 (11, 32)**		**19 (7, 31)**
10. Suitable for publication	22	3.6 (1.1)	22	4.0 (0.8)	22	3.5 (1.1)	22	3.0 (1.0)	22	2.9 (1.2)	22	2.0 (1.0)
		4 (2, 5)		4 (3, 5)		4 (1, 5)		3 (1, 4)		3 (1, 5)		2 (1, 4)
11. Suitable for final report	22	4.0 (1.0)	22	4.2 (0.7)	22	3.6 (1.1)	22	3.2 (1.1)	22	3.0 (1.2)	22	2.4 (1.0)
		4 (2, 5)		4 (3, 5)		4 (1, 5)		4 (1, 5)		3 (1, 5)		2 (1, 4)
12. Suitable for interim analysis	22	4.0 (1.0)	22	4.2 (0.8)	22	3.3 (1.1)	22	3.0 (1.0)	22	2.8 (1.4)	22	2.9 (1.3)
		4 (2, 5)		4 (2, 5)		3 (1, 5)		3 (1, 5)		3 (1, 5)		3 (1, 5)
13.Exploratory analysis	22	4.0 (0.8)	23	4.0 (1.0)	23	3.6 (0.9)	23	3.5 (0.9)	23	3.2 (1.2)	23	3.3 (1.3)
		4 (2, 5)		4 (1, 5)		4 (1, 5)		4 (1, 5)		3 (1, 5)		4 (1, 5)
14.Explanatory analysis	22	4.0 (1.0)	23	4.2 (0.7)	23	3.5 (1.0)	23	3.0 (1.0)	23	2.7 (1.2)	23	2.3 (1.1)
		4 (2, 5)		4 (3, 5)		4(1, 5)		3 (1, 5)		3 (1, 5)		2 (1, 4)
Ranking	**20**	**2.5 (1.5)**	**20**	**1.9 (1.1)**	**19**	**2.9 (1.4)**	**20**	**4.3 (1.6)**	**16**	**4.3 (1.8)**	**18**	**4.8 (1.4)**
		2 (1, 6)		**2 (1, 5)**		**3 (1, 5)**		**5 (1, 6)**		**5 (1, 7)**		**5 (2, 7)**

*** Overall score is the sum total of questions 1-7**

Supplement 5: Table and figures summarising initial appraisals of all plots by outcome type

Figure A.32: Single time-to-event outcomes



Box plot of overall scores ordered by highest to lowest mean values (higher scores indicate better performance). **b. Box plot of rankings** ordered by best to worst mean rank (lower ranking indicates preferred plot).
Note: X indicates median values.

Supplement 5: Table and figures summarising initial appraisals of all plots by outcome type

Table A.6: Plots suitable for **Multiple Continuous Outcomes** – summary of scores

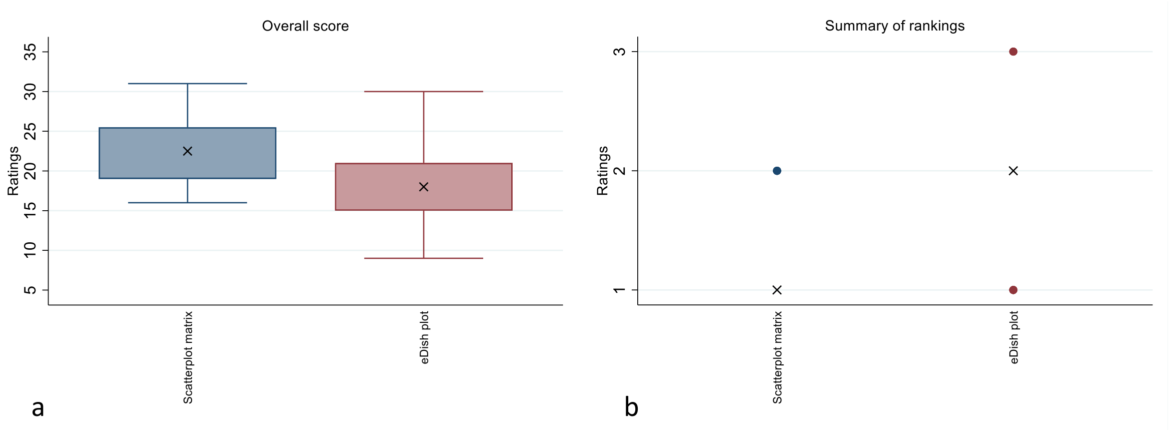
Question	**Scatterplot matrix**		**E-dish**	
	n	Mean (SD)	n	Mean (SD)
		Median (Min, Max)		Median (Min, Max)
1.Effect size	22	2.5 (1.1)	20	1.9 (0.9)
		2 (1, 5)		2 (1, 4)
2.Direction of effect	22	3.0 (1.1)	20	2.5 (1.1)
		3 (1, 5)		2 (1, 5)
3.Uncertainty	22	1.7 (0.8)	20	1.6 (0.8)
		2 (1, 3)		1 (1, 3)
4.Supplementary data needed	22	3.2 (1.1)	20	2.4 (1.1)
		3 (1, 5)		2 (1, 5)
5.Understandable	22	4.5 (0.8)	20	3.8 (1.3)
		5 (2, 5)		4 (1, 5)
6.Understandable non-stats	22	4.3 (0.6)	20	3.6 (1.0)
		4 (3, 5)		4 (1, 5)
7.Multi-arm studies	20	2.8 (1.4)	18	2.2 (1.3)
		3 (1, 5)		2 (1, 5)
8.Limits numbers	22	2.9 (1.2)	19	2.1 (1.2)
		3 (1, 5)		2 (1, 5)
9.Overall score	**20**	**22.8 (4.3)**	18	18.6 (5.4)
		23 (16, 31)		18 (9, 30)
10. Suitable for publication	20	2.8 (1.3)	19	2.7 (1.2)
		3 (1, 5)		3 (1, 5)
11. Suitable for final report	20	3.4 (1.3)	19	3.2 (1.2)
		4 (1, 5)		3 (1, 5)
12. Suitable for interim analysis	20	4.0 (1.0)	19	3.4 (1.2)
		4 (1, 5)		4 (1, 5)
13.Exploratory analysis	20	4.2 (0.8)	19	3.7 (1.1)
		4 (3, 5)		4 (1, 5)
14.Explanatory analysis	20	3.0 (1.1)	19	2.7 (1.2)
		3 (1, 5)		3 (1, 5)
Ranking	**18**	**1.2 (0.4)**	**18**	**1.9 (0.6)**
		1 (1, 2)		**2 (1, 3)**

* Overall score is the sum total of questions 1-7

Excludes summary for vector plots as deemed not applicable during the meeting.

Supplement 5: Table and figures summarising initial appraisals of all plots by outcome type

Figure A.33: Multiple continuous outcomes



Box plot of overall scores ordered by highest to lowest mean values (higher scores indicate better performance). **b. Box plot of rankings** ordered by best to worst mean rank (lower ranking indicates preferred plot).

Note: X indicates median values. Excludes summary for vector plots.

Supplement 5: Table and figures summarising initial appraisals of all plots by outcome type

Table A.7a: Plots suitable for **Single Continuous Outcomes** – summary of scores

Question	**Empirical distribution of max change**		**Histogram of max change**		**Delta plot**		**Line graph - change**		**Boxplot - change**		**Violin plot - change**	
	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)
		Median (Min, Max)		Median (Min, Max)		Median (Min, Max)		Median (Min, Max)		Median (Min, Max)		Median (Min, Max)
1.Effect size	22	2.2 (1.2)	22	2.5 (1.2)	23	1.7 (1.1)	22	3.4 (1.4)	22	2.7 (1.3)	22	2.5 (1.3)
		2 (1, 4)		2 (1, 5)		1 (1, 5)		4 (1, 5)		3 (1, 5)		2 (1, 5)
2.Direction of effect	22	3.1 (1.2)	22	3.0 (1.1)	23	1.6 (0.8)	22	4.0 (0.9)	22	3.1 (1.1)	22	3.0 (1.1)
		3 (1, 5)		3 (1, 4)		1 (1, 4)		4 (2, 5)		3 (1, 5)		3 (1, 5)
3.Uncertainty	22	1.3 (0.5)	22	1.7 (1.0)	23	1.2 (0.4)	22	3.9 (0.9)	22	3.3 (1.3)	22	3.0 (1.1)
		1 (1, 2)		1 (1, 4)		1 (1, 2)		4 (2, 5)		4 (1, 5)		3 (1, 5)
4.Supplementary data needed	22	2.5 (1.1)	22	3.5 (0.9)	23	1.7 (0.8)	22	3.6 (1.0)	22	3.4 (0.9)	22	3.1 (0.8)
		3 (1, 4)		4 (2, 5)		2 (1, 3)		4 (2, 5)		4 (2, 5)		3 (2, 4)
5.Understandable	21	3.0 (0.9)	22	4.5 (0.7)	23	1.7 (0.8)	22	4.4 (0.7)	22	4.5 (0.7)	22	3.8 (0.7)
		3 (1, 4)		5 (3, 5)		2 (1, 4)		4 (3, 5)		5 (3, 5)		4 (3, 5)
6.Understandable non-stats	22	2.4 (0.9)	22	4.2 (0.6)	23	1.4 (0.6)	22	4.1 (0.8)	22	3.9 (0.8)	22	2.9 (0.9)
		3 (1, 4)		4 (3, 5)		1 (1, 3)		4 (3, 5)		4 (2, 5)		3 (2, 5)
7.Multi-arm studies	21	3.9 (0.9)	20	3.2 (0.9)	21	2.4 (1.1)	20	4.0 (0.6)	20	4.0 (0.8)	19	3.7 (0.6)
		4 (1, 5)		3 (2, 5)		2 (1, 4)		4 (3, 5)		4 (2, 5)		4 (3, 5)
8.Limits numbers	22	2.5 (1.6)	20	2.4 (1.5)	22	2.0 (1.2)	22	2.6 (1.4)	22	2.7 (1.4)	22	2.7 (1.4)
		3 (1, 5)		2 (1, 5)		2 (1, 5)		3 (1, 5)		3 (1, 5)		3 (1, 5)
9.Overall score	**22**	**17.7 (5.9)**	**21**	**21.9 (5.8)**	**21**	**12.0 (3.8)**	**21**	**26.3 (7.3)**	**21**	**23.7 (6.3)**	**21**	**20.3 (5.4)**
		19 (0, 26)		**21 (0, 28)**		**11 (7, 20)**		**28 (0, 35)**		**25 (0, 34)**		**21 (0, 26)**
10. Suitable for publication	22	2.9 (1.2)	21	3.2 (1.0)	21	1.6 (0.6)	21	4.0 (0.7)	21	3.6 (1.0)	21	3.4 (0.9)
		3 (1, 5)		3 (1, 5)		2 (1, 3)		4 (3, 5)		4 (1, 5)		3 (1, 5)
11. Suitable for final report	22	3.1 (1.1)	21	3.7 (1.0)	21	1.8 (0.9)	21	4.0 (0.7)	21	3.6 (1.1)	21	3.4 (1.0)
		4 (1, 5)		4 (1, 5)		2 (1, 4)		4 (3, 5)		4 (1, 5)		3 (1, 5)
12. Suitable for interim analysis	22	3.3 (1.2)	21	3.6 (1.1)	21	2.0 (1.1)	21	3.9 (0.8)	21	3.8 (1.0)	21	3.5 (0.9)
		4 (1, 5)		4 (1, 5)		2 (1, 5)		4 (3, 5)		4 (1, 5)		3 (1, 5)
13.Exploratory analysis	22	3.6 (1.1)	21	3.7 (0.9)	21	2.4 (1.1)	21	4.0 (0.8)	21	3.9 (1.0)	21	3.7 (0.7)
		4 (1, 5)		4 (2, 5)		2 (1, 5)		4 (2, 5)		4 (1, 5)		4 (2, 5)
14.Explanatory analysis	22	2.9 (1.2)	21	3.4 (1.1)	21	1.8 (0.8)	21	4.0 (0.7)	21	3.7 (1.1)	21	3.5 (0.8)
		3 (1, 5)		3 (2, 5)		2 (1, 3)		4 (2, 5)		4 (1, 5)		3 (2, 5)
Ranking	**17**	**6.8 (1.9)**	**17**	**4.9 (2.1)**	**19**	**8.1 (1.8)**	**20**	**2.0 (1.2)**	**18**	**3.6 (2.1)**	**18**	**3.8 (1.8)**
		8 (3, 9)		**5 (1, 8)**		**9 (2, 9)**		**2 (1, 4)**		**4 (1, 7)**		**4 (1, 7)**

* Overall score is the sum total of questions 1-7

Supplement 5: Table and figures summarising initial appraisals of all plots by outcome type

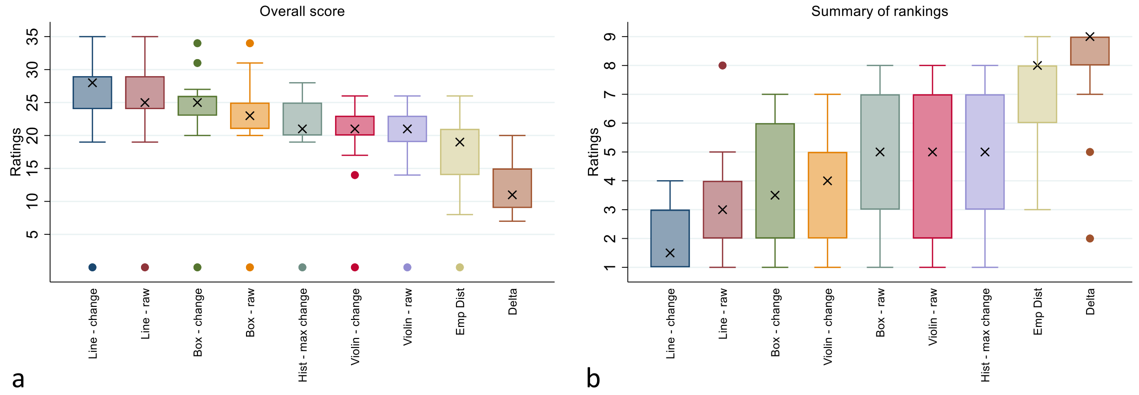
Table A.7b: Plots suitable for **Single Continuous Outcomes** – summary of scores

Question	**Line graph - raw**		**Box plot - raw**		**Violin - raw**	
	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)
		Median (Min, Max)		Median (Min, Max)		Median (Min, Max)
1.Effect size	22	3.0 (1.4)	22	2.6 (1.3)	22	2.4 (1.3)
		3 (1, 5)		2 (1, 5)		2 (1, 5)
2.Direction of effect	22	3.7 (1.1)	22	3.0 (1.0)	22	2.9 (1.0)
		4 (1, 5)		3 (1, 5)		3 (1, 4)
3.Uncertainty	22	3.6 (1.0)	22	3.3 (1.3)	22	3.0 (1.1)
		4 (1, 5)		4 (1, 5)		3 (1, 5)
4.Supplementary data needed	22	3.4 (1.1)	22	3.2 (1.0)	22	3.0 (1.0)
		3 (2, 5)		3 (2, 5)		3 (1, 5)
5.Understandable	22	4.5 (0.6)	22	4.4 (0.8)	22	3.8 (0.9)
		5 (3, 5)		5 (2, 5)		4 (2, 5)
6.Understandable non-stats	22	4.2 (0.7)	22	3.9 (0.9)	22	2.9 (1.0)
		4 (3, 5)		4 (2, 5)		3 (1, 5)
7.Multi-arm studies	20	4.0 (0.6)	20	3.9 (0.7)	19	3.6 (0.6)
		4 (3, 5)		4 (2, 5)		4 (3, 5)
8.Limits numbers	22	2.6 (1.4)	22	2.7 (1.4)	22	2.6 (1.4)
		2 (1, 5)		3 (1, 5)		3 (1, 5)
9.Overall score	**21**	**25.3 (7.4)**	**21**	**23.1 (6.4)**	**21**	**20.1 (5.4)**
		25 (0, 35)		**23 (0, 34)**		**21 (0, 26)**
10. Suitable for publication	21	3.8 (1.0)	21	3.4 (1.1)	21	3.2 (1.0)
		4 (1, 5)		3 (1, 5)		3 (1, 5)
11. Suitable for final report	21	3.9 (1.0)	21	3.6 (1.0)	21	3.4 (1.0)
		4 (1, 5)		4 (1, 5)		3 (1, 5)
12. Suitable for interim analysis	21	3.8 (1.0)	21	3.6 (1.0)	21	3.3 (1.0)
		4 (1, 5)		4 (1, 5)		3 (1, 5)
13.Exploratory analysis	21	3.9 (0.8)	21	3.8 (0.9)	20	3.6 (0.7)
		4 (2, 5)		4 (1, 5)		4 (2, 5)
14.Explanatory analysis	21	3.8 (1.0)	21	3.5 (1.1)	20	3.3 (0.9)
		4 (2, 5)		3 (1, 5)		3 (2, 5)
Ranking	**18**	**3.2 (1.8)**	**18**	**4.7 (2.2)**	**18**	**4.7 (2.3)**
		3 (1, 8)		**5 (1, 8)**		**5 (1, 8)**

* Overall score is the sum total of questions 1-7

Supplement 5: Table and figures summarising initial appraisals of all plots by outcome type

Figure A.34: Single continuous outcomes



a. Box plot of overall scores ordered by highest to lowest mean values (higher scores indicate better performance). **b. Box plot of rankings ordered** by best to worst mean rank (lower ranking indicates preferred plot).
Note: X indicates median values.

Supplement 6: Tables summarising Mentimeter votes to decide which plots to take forward and amendments

Table A.8: Decisions for **Multiple Binary Outcomes**

Question		n	%
Should we **keep** the dot plot?	Yes	19	100
	No	0	0
Should we **keep** the stacked bar chart?	Yes	18	95
	No	1	5
Should we **keep** the bar chart?	Yes	17	81
	No	4	19
Should we **exclude** the tendril plot?	Yes	20	100
	No	0	0
Should we **exclude** the alluvial plot?	Yes	17	94
	No	1	6
Should we **exclude** alternative 3?	Yes	18	100
	No	0	0
Should we **keep** the volcano plot?	Yes	11	65
	No	6	35
Should we **keep** alternative 1?	Yes	8	47
	No	9	53
Should we **keep** alternative 2?	Yes	10	50
	No	10	50
Should we keep the **heat** map?	Yes	3	16
	No	16	84
Should we keep the **star** plot?	Yes	2	10
	No	18	90

Supplement 6: Tables summarising Mentimeter votes to decide which plots to take forward and amendments

Table A.9: Decisions about amendments for the recommend plots for **Multiple Binary Outcomes**

		n	%
Are we happy to recommend the dot plot as it is unedited?	Yes	14	67
	No	7	33
Do we want to add in counts and number of participants into the data table?	Yes	12	60
	No	8	40
Are we happy to recommend the stacked bar chart as it is unedited?	Yes	16	80
	No	4	20
Do we want to recommend the volcano in light of possible alternative?	Yes	7	35
	No	13	65
	Yes	7	35
Do we want to recommend the alternative 2 instead?	No	13	65

Table A.10: Decisions about plots in the **Single Binary Outcome** setting

		n	%
Is it helpful to use a plot in this setting?	Yes	14	67
	No	7	33
Would you like to see this in bar chart?	Yes	15	75
	No	5	25
Would you prefer the data to be presented by bars or dots?	Bars	15	79
	Dots	4	21
Should we present as two separate charts one above the other aligned vertically?	Yes	9	50
	No	9	50
Should we present as two separate charts one above the other aligned vertically?	Context specific e.g. only 2 arms then horizontal, >2 arms then stacked	11	58
	Stacked one above the other	7	37
	Horizontal - side by side	1	5

Supplement 6: Tables summarising Mentimeter votes to decide which plots to take forward and amendments

Table A.11: Decisions for plots to recommend in the **Single Time-to-Event** setting

		n	%
Should we recommend KM or Cumulative Hazard plots?	Cumulative hazard	3	17
	Kaplan-Meier	15	83
What should the table at the bottom of the KM plot contain?	No table	1	6
	Minimum - at risk table only (by arm)	4	24
	Full table as per KMUNICATE	12	71
Should we recommend the survival ratio plot as an alternative to the KM?	Yes	12	67
	No	6	33
Should we recommend the MCF plot for displaying information on repeated events?	Yes	15	88
	No	2	12
Should the table at the bottom of the MCF plot only contain the number at risk (by arm)	Yes	17	94
	No	1	6

Acronyms: KM – Kaplan-Meier; MCF – Mean Cumulative Function

Table A.12: Decisions for plots in the **Multiple Time-to-Event** setting

		n	%
Should we recommend any of these plots?	Matrix of multiple KM	8	40
	Bar chart of median time-to-event	0	0
	Heat map/alternative 4	2	10
	None of these	10	50

Acronyms: KM – Kaplan-Meier

Supplement 6: Tables summarising Mentimeter votes to decide which plots to take forward and amendments

Table A.13: Decisions for plots in the **Single Continuous Outcome** setting

		n	%
Should we recommend a version of the line chart?	Yes	17	94
	No	1	6
Should we recommend a version of the boxplot?	Yes	9	53
	No	8	47
Should we recommend a version of the violin plot?	Yes	12	67
	No	6	33
Should we recommend a version of the histogram?	Kernel density	11	61
	Histogram	0	0
	Neither	7	39

Table A.14: Decisions for plots in the **Multiple Continuous Outcome** setting

		n	%
Should we recommend the scatterplot matrix?	Yes	16	94
	No	1	6

Supplement 7: Free text comments accompanying initial appraisals of recommended plots

1. *Multiple binary outcomes*

    i. *Dot plot*

Proposals to the dot plot included adding numerical raw data via either a data table on the right-hand side of the plot or labelling data points on the left-hand side of the plot in order to enrich information presented and to provide an alternative to the typical frequency tables presented in publications. Concerns were raised about the inclusion of confidence intervals in this plot as this could encourage use as a proxy for hypothesis tests but discussions indicated that this could be caveated by including a caution to avoid such interpretation in the recommendations for use.

    ii. *Stacked bar charts*

Preference was for stacked bar charts of percentages with at least one event and inclusion of bar labels of frequencies or counts of events. Imposing a meaning to the order of bars was also advocated.

2. *Single binary outcomes*

    i. *Bar chart of counts*

A boxplot or dot plot of a summary measure of count data was suggested to replace the bar chart as a means to summarise count data, however, there was not whole group support for this idea. There was variation in preferences for layout of the bar chart with some preferring side-by-side plots for each treatment group and others preferring plots stacked one above the other for each treatment group. Discussions highlighted that some participants question the need for this plot, for example, with one participant commenting, "*is aggregation of data like this helpful?*", and others felt there could be difficulty in interpreting these plots.

Supplement 7: Free text comments accompanying initial appraisals of recommended plots

Discussions concluded that this plot might only be useful for summaries of serious events or pre-specified events.

3. *Single time-to-event outcomes*

    i. *Kaplan-Meier*

Amendments discussed for the Kaplan-Meier plot included: incorporating an extended at risk table including the number of participants that remain ‘at risk’, the cumulative number that have been censored and the cumulative number that have experienced an event at each discrete time point; providing a clear definition of what ‘survival’ means in the context of analysing harm outcomes in the recommendations; and incorporating a between group comparison. This latter point prompted discussions and a suggestion to consider survival ratio plots proposed by Newell et al.[245] Survival ratio plots were not formally considered via appraisals but were incorporated into discussions for consideration. Given the context of use we instead refer to the survival ratio plot as the event-free ratio plots. Discussions revealed some concerns about use of time-to-event plots in this setting and concluded that recommendations should caution users to bear in mind the consequences of competing risks.

    ii. *Mean cumulative function*

Proposals for the mean cumulative function included adding confidence interval bands and at risk tables. Discussions covered whether grouping all events together in this plot should be encouraged or that instead the recommendation should be for use when analysing pre-specified events of interest. Participants endorsed the latter, recommending use to account for recurrent events. Discussions also indicated that recommendation should include clear text descriptions explaining the interpretation of this plot given its novelty and clarifying that it adequately accounts for censoring.

Supplement 7: Free text comments accompanying initial appraisals of recommended plots

4. *Multiple time-to-event outcomes*

Discussions on the plots for displaying multiple time-to-event outcomes highlighted the lack of a suitable plot for consideration and that further development work in this area is needed. However, in the interim, initial discussions indicated that the matrix of Kaplan-Meier plots could be utilised and was taken forward for further critique, although ultimately was not recommended.

i. Matrix of Kaplan-Meier

Discussions indicated that the matrix of Kaplan-Meier plots required incorporation of confidence bands and tables of numbers at risk as per the individual Kaplan-Meier plots. Participants indicated that this plot would be useful to detect disproportionalities for pre-specified events but that the number of events looked at would need to be limited to be useful. To avoid encouraging performance of many hypothesis tests it was also highlighted that it should be clearly specified that this plot should be used as a way to display risk over time to help identify disproportionalities and raise signals for adverse drug reactions. Alternatives that incorporate information on recurrent events are still needed.

5. *Multiple continuous outcomes*

i. Scatterplot matrix

Discussed amendments to the scatterplot matrix included ways to help ease the problems created by overlapping points, inclusion of reference lines and labels for outliers.

Supplement 7: Free text comments accompanying initial appraisals of recommended plots

6. Single continuous outcomes

i. Line chart

Discussions focused on the appropriate statistic to display on the line chart as well as advocating for inclusion of tables with numbers at risk at the bottom of the plot that are typically seen on time-to-event plots such as the Kaplan-Meier plot.

ii. Violin plot

A proposal to remove the duplication of information in the 'mirrored' distributions of the violin plot was discussed and an indication of a preference for violin plots over box plots was voiced.

iii. Histogram/Kernel density plot

Discussions indicated that participants wished to see this information presented graphically but would prefer to see it displayed in kernel density plots instead of histograms, which would overcome the problems of overlap encountered in the histogram.