Title: Visualising harms in Randomised Controlled Trial publications: a consensus and recommendations

Rachel Phillips*[1, 2], Senior Lecturer in Medical Statistics

Suzie Cro[1], Advanced Research Fellow in Medical Statistics

Graham Wheeler[1], Deputy Head of Statistics and Senior Lecturer in Clinical Trials Statistics

Simon Bond[3], Senior Statistician

Tim P. Morris[4], Senior Research Associate in Medical Statistics

Siobhan Creanor[5], Director of Exeter CTU and Professor of Medical Statistics & Clinical Trials

Catherine Hewitt[6], Deputy Director of York Trials Unit and Professor of Trials and Statistics

Sharon Love[4], Associate Professor

Andre Lopes[7], Senior Research Fellow in Medical Statistics

Iryna Schlackow[8], Senior Researcher in Health Economics

Carrol Gamble[9], Director of Liverpool Clinical Trials Centre and Professor of Medical Statistics

Graeme MacLennan[10], Director of the Centre for Health Care Randomised Trials and Professor of Medical Statistics

Chris Habron[11], Expert Statistical Scientist

Anthony C Gordon[12], Chair in Anaesthesia and Critical Care and Consultant in Intensive Care Medicine

Nikhil Vergis[13], Clinical Academic and Consultant in Gastroenterology and General Internal Medicine

Tianjing Li[14], Associate Professor

Riaz Qureshi[14], Assistant Professor

Colin C Everett[15], Senior Medical Statistician

Jane Holmes[16], Senior Medical Statistician

Amanda Kirkham[17], Senior Biostatistician

Clare Peckitt[18], Royal Marsden CTU Manager (Statistics)

Sarah Pirrie[17], Principal Biostatistician

Norin Ahmed[19], Research Fellow in Medical Statistics

Laura Collett[20], Medical Statistician

Victoria Cornelius[1], Head of Statistics and Trial Methodology at Imperial Clinical Trials Unit and Reader in Medical Statistics

[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

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

[4]MRC Clinical Trials Unit at UCL, 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, UCL, 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 Product Ltd, 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, Colorado, 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, The Royal Marsden NHS Foundation Trust, London, UK

[19]Comprehensive Clinical Trials Unit, UCL, London, UK

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

*Corresponding author

Email: r.phillips@imperial.ac.uk

Address: Imperial Clinical Trials Unit, Imperial College London, 1st Floor Stadium House, 68 Wood Lane, London, W12 7RH

ORCID ID: https://orcid.org/0000-0002-3634-7845

Twitter: @RachPips

ABSTRACT

Objective: To improve communication of harm in RCT publications we identified researchers' recommendations for visualising harm outcomes.

Design: Consensus study evaluating visualisation methods.

Setting: 15 UKCRC registered CTUs, an academic population health department, Roche Product Ltd and the BMJ.

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

Data sources: Visualisations were primarily identified via a methodological review of statistical methods developed specifically to analyse harm outcomes, these were considered alongside visualisations recommended by consensus group members.

Interventions: None

Main outcomes measured: Consensus for visualisations to recommend achieved over a series of three meetings with participants. Participants reviewed and critically appraised candidate visualisations against an agreed framework. Appraisals were summarised and presented back to participants to inform discussions. After discussions participants voted on whether to endorse each visualisation.

Eligibility criteria: Visualisation receiving at least 60% of the available votes were endorsed. Scores marginally below this threshold (50-60%) were revisited for further discussions and votes retaken until a consensus was reached.

Results: Twenty-eight visualisations were considered, of which ten are recommended to researchers to consider in publications of main research findings. The choice of visualisations to present will depend on outcome type e.g., binary, count, time-to-event or continuous and the scenario e.g., summarising multiple emerging events or one event of interest. A decision tree to assist trialists decide which visualisations to use is presented.

Examples of each endorsed visualisation, along with example interpretation, potential limitations and signposting to code for implementation across a range of standard statistical software are provided. 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 harm information and thus enable informative interpretation, especially valuable for assessing the profile of harm. Whilst we endorse each of the visualisations presented, we also note their limitations and provide examples of where their use would be inappropriate. Though 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. We recommend trialists continue to examine crude numbers alongside visualisations to fully understand harm profiles.

PRINT ABSTRACT

Study question: To identify researchers' recommendations for visualising harm outcomes in RCT publications.

Methods: A series of consensus meetings was held comprising of 20 statisticians from 15 UKCRC registered CTUs, an academic health economist, an industry statistician, and a data graphics designer from the BMJ. Visualisations were primarily identified via a methodological review of statistical methods developed specifically to analyse harm outcomes. Participants reviewed and critically appraised candidate visualisations against an agreed framework. Appraisals were summarised and presented back to participants to inform discussions. After discussions participants voted on whether to endorse each visualisation. A threshold of 60% was used to indicate endorsement. Scores marginally below this threshold (50-60%) were revisited until a consensus could be reached. Clinician feedback was incorporated into the explanatory information provided in the recommendations to aid understanding and interpretation.

Study answer and limitations: Twenty-eight visualisations were considered, of which ten are recommended to researchers to consider in publications of main research findings. The choice of visualisations to present will depend on outcome type e.g., binary, count, time-to-event or continuous and the scenario e.g., summarising multiple emerging events or one event of interest. A decision tree to help trialists decide which visualisations to use is presented, however the statistician and clinical trial team must ultimately decide the most appropriate visualisations for their data and objectives. Examples of each endorsed visualisation, along with example interpretation, potential limitations and signposting to code for implementation across a range of standard statistical software are provided.

What this study adds: Researchers have called for guidance on appropriate methods for the analysis of harm outcomes and case studies detailing examples of use. We address this by providing recommendations and tools to help researchers decide which visualisations to

use. Increasing the use of visualisations for harm outcomes in clinical trial manuscripts will provide clearer presentation of harm information and thus enable more informative interpretation, especially valuable for assessing the profile of harm.

Funding, competing interests, and data sharing: RP was funded by NIHR grant DRF-2017-10-131 to undertake this work. SC is supported by NIHR grant NIHR300593 and TPM was supported MRC grants MC_UU_00004/07, MC_UU_12023/21 and MC_UU_12023/29. All authors declare no conflicts of interest. The datasets used in this analysis are available from GlaxoSmithKline via ClinicalStudyDataRequest.com and the synthetic dataset example is available for download via associated Stata packages.

Print figure: Figure 1

SUMMARY BOX

What is already known on this topic?

RCTs provide a valuable source of data to compare harm outcomes between treatment groups and can help identify potential signals for adverse (drug) reactions, but there is evidence of suboptimal practices when reporting data on harm outcomes in clinical trial manuscripts. Harm outcomes data is complex but visualisations can provide a clear summary of the harm profile and help identify potential adverse (drug) reactions. Researchers have requested guidance on appropriate visualisations for harm outcomes and case studies detailing examples of use.

What this study adds

We undertook a consensus and endorsed visualisations to communicate harms in the RCT setting that can be used as an alternative to the widely used contingency tables are presented alongside a decision tree to aid researchers in their choice of visualisations. The choice of visualisation will depend on the outcome type (e.g., binary, count, time-to-event or continuous), the scenario (e.g., summarising multiple emerging events or one event of interest), the design of the trial (trials with more than two treatment groups require more care) and the purpose of the plot (e.g., to communicate information about the entire harm profile or convey a direct message about a particular event of interest). Increasing the use of visualisations for harm outcomes in clinical trial manuscripts and reports will provide clearer presentation of harm information and thus enable informative interpretation, especially valuable for assessing the profile of harm.

INTRODUCTION

A well-designed graphic is an effective way to communicate a message to a range of audiences and help identify patterns in data that might otherwise be missed.[1] In 1983 Tufte stated, "*of all methods for analyzing 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 (i.e. non-prespecified events that are reported and collected during the trial and may be unexpected) where there are a lot of complex data, visualisations can help summarise harm profiles and identify potential adverse (drug) reactions (A(D)Rs). Trials can also prespecify events as harm outcomes of interest to follow. These may be known or suspected to be associated to the intervention, or followed for reasons of interest and visualisations have much to offer here too. Trial reporting guidelines encourage the use of visualisations for exploring harm outcomes including: the CONSORT extension to harms, the 2016 recommendations to improve adverse event (AE) reporting from industry representatives and journal editors, a pharmaceutical industry standard from the Safety Planning, Evaluation and Reporting Team (SPERT) and guidance from regulators on statistical principles in clinical trials (ICH E9).[3-6] There are an abundance of potential visualisations available but their use in journal articles is limited.[3 4 7 8] A systematic review performed in 2018 found that only 12% of journal articles made use of visual summaries for AE data, a finding supported by a 2019 survey of UK Clinical Research Collaboration (UKCRC) Clinical Trial Unit (CTU) statisticians.[9 10] However, a 2016 survey of industry statisticians suggested that in-house practice in this sector might differ.[11] Evidence suggests the current practice in journal articles is to summarise harm outcomes from randomised controlled trials (RCTs) in simple tables of frequencies and percentages, despite the advantages visualisations offer.[12] Key terms and definitions relating to harm outcomes used throughout this article are presented in table 1.[13 14]

Advances in computer software have improved trialists' capability to produce visualisations, but there is lack of guidance 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*".[10, 15] Therefore, with a range of visualisation options available and the increasing ease in which they can be implemented, we sought a consensus to support researchers in their choice of visualisations for RCT publications. In this article, in collaboration with the UKCRC CTU Statistics Operations Group, we provide recommendations on which visualisations researchers should consider including in the publication of their main research findings.

METHODS

A series of consensus meetings was held comprising of 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. All of whom are experienced clinical trialists and/or have an interest in visualisations. The group reviewed and critically evaluated (against an agreed framework) 28 plots proposed for visualising data on harm outcomes and refined these plots as necessary, predominantly focusing on clinical trials of an investigational medicinal product (CTIMP). Examples of each of the candidate plots was produced using data from one of four completed parallel arm pharmacological RCTs and a synthetic dataset (further details on each and instructions for access are available in supplement 1). 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 RCTs that make such inferences. The methods detailing the identification of the considered plots, consensus process and how the

recommendations were developed are provided in supplement 1. In the following each of the endorsed plots is described, example interpretation is given, and we then provide our recommendation. Thumbnail images are included in the paper for each plot but when reading interpretations the full sized plot in the supplement 2 will be useful for reference.

RESULTS

Endorsed visualisations that researchers should consider including in the publication of their main research findings according to harm outcome type and number of events (either single outcomes or multiple outcomes simultaneously) are displayed as thumbnail images in figure 1 and supplement 2 figures A.1-A.10 and are discussed in detail below. Outcome type include 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. Endorsed visualisations, according to whether they assess the entire profile or convey a direct message about a particular event(s) of interest, are presented alongside the recommendations for use in table 2. To help trialists decide which visualisation to use, a decision tree (figure 2) and a summary table of required outcome characteristics (table 3) 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. Visualisations (n=18) that were considered but not endorsed are included for information in supplement 3 with descriptions and the potential adaptations discussed (figures A.11-A.28).

Recommendations for multiple binary outcomes

1. **Dot plot**

Plot description

The dot plot summarises both the absolute and relative risk for multiple events (figure 1 image 1, supplement 2 figure A.1). The left panel displays the percentage of participants experiencing an event (labelled on the vertical axis) in each treatment group. The central panel displays a measure of comparison, in this example the relative risk of observing each event in the treatment group compared to 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 from bottom to top by increasing relative risk. 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 the example (figure 1 image 1, supplement 2 figure A.1) the overall impression is that point estimates for the relative summary statistic are evenly distributed on either side of the vertical line of no difference (relative risk = 1) but with great differences in levels of precision 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 experiencing this event. There is also evidence of 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 experiencing 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 one, whilst these estimates look small in comparison to the other relative risks, it is clear to see from the left hand side of the plot the marked difference in absolute numbers and from the data table the

large numbers experiencing these events. This suggests a potential beneficial effect of the intervention on these harm outcomes that may warrant closer inspection.

Recommendation

The group unanimously endorsed the dot plot for presenting data on 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 (endorsement levels for the other recommended plots can be found in supplement 6).

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 may also prefer to present the data table in the central panel so that it appears alongside the absolute summary. It is possible to create this plot in grayscale without loss of meaning. Whilst it may be possible to add in a limited number of additional arms for multi-arm studies through incorporation of multiple non-overlapping estimates on the same plot, for example, using jittering, it may soon become incomprehensible as the number of active treatment groups increases.

Limitations

Confidence intervals around the relative differences are useful to identify potential signals of harm for further investigation, but they should not be used as a proxy for hypothesis testing, which will increase the chance of finding spurious significant differences due to multiple hypothesis tests performed.[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, where events are rare it may not be appropriate to present the entire 0 to 100 scale for the absolute summary. When presenting the odds ratios or risk ratios, if there are zero events in one of the treatment

groups a common, simple, correction is to add half an event to each group (numerator and denominator). This is a commonly used continuity correction but has been shown to be inferior when undertaking meta-analyses on rare events, therefore alternative corrections may warrant consideration.[17 18] Whilst this plot gives a comprehensive overview, some potentially important pieces of information are not included such as information on relative severity of different harm outcomes and whilst recurrent events within participants can be presented via the incident rate ratio, there is not an easy way to display this information on the left panel. In scenarios where it is important to display information on severity, the stacked bar chart can be used (see section: *2. Stacked bar chart*) and for recurrent events the mean cumulative function plot can be used (see section: *5. Mean cumulative function plot)*.

Software

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

2. 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 i.e. if a participant had the same event twice, once classified as mild and once as moderate this participant would be counted once as experiencing a moderate event (figure 1 image 2, supplement 2 figure A.2). 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 experiencing 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 the example (figure 1 image 2, supplement 2 figure A.2) the most frequent events experienced by participants at least once are blood and lymphatic events and gastrointestinal disorders. Whist there were more blood and lymphatic events in the placebo group, the stacked bar chart reveals that the numbers 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 numbers experiencing mild events. For gastrointestinal disorders, the stacked bar chart revealed that there were fewer events in the intervention group across each of the severity grades in comparison to the placebo group. The plot also revealed that events classified as 'other' were dominated by severe and moderate events in the intervention group compared to the placebo group, which could warrant closer inspection of what these events were. In contrast to the dot plot, the stacked bar chart highlights the most frequent events, due to the increased physical space these events occupy, whereas in the dot plot the most frequent events take up the least space in the central panel due to 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. It can be used to informally compare severe or severe plus moderate events or the overall number of events between groups. It is recommended that treatment groups are 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 multi-arm studies and graduation in colour from black to white is possible to avoid use of colour. This plot could be adapted to the single event setting by replacing events on the vertical axis with some representation of time e.g. visits or treatment cycle, an example of which can be found in Thanarajasingam et al.[20]

Limitations

Direct comparisons within stacked bars 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. This plot promotes presenting 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, there is no explicit display of the effect sizes for differences between groups.

Software

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

Recommendations for single binary outcomes

3. **Bar chart – for counts**

Plot description

A bar chart to present information on the number of events or the count of events experienced per participant (figure 1 thumbnails 3a and 3b, supplement 2 figures A.3a and A.3b). Each bar represents the percentage of participants with 0, 1, 2 etc. events for each treatment group.

Implementation and interpretation

Figure 1 image 3a (supplement 2 figure A.3a) displays the distribution for the multiple events experienced by participants, with placebo participants experiencing higher numbers of multiple events more often. In figure 1 image 3b (supplement 2 figure A.3b), the distributions indicate that participants in either of the intervention groups experience multiple events more often compared to the placebo group.

Recommendations

The bar chart is recommended to present information on the number of events experienced. This is a simple plot that 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. It can be used to present information on an overall summary of events such as the total number of serious adverse events or for a limited number or single events of interest. It 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 (figure 1 image 3a) when comparing two treatment groups. When there are more than two treatment groups, separate plots stacked above each other for each group (figure 1 image 3b) is the recommended alternative.

Potential amendments

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

Limitations

Whilst 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 contributing to it. 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, whilst it 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 (Stata, `graph bar`; R, `barplot` or the `ggplot2` package with `geom_bar`; SAS, `proc gchart`).

Recommendations for single time-to-event outcomes

4. **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 (figure 1 image 4, supplement 2 figure A.4). 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 that remain ‘at risk’ for the specific event of interest, the cumulative number that have been censored and the cumulative number that have experienced the event of interest at each discrete time point.

Implementation and interpretation

In figure 1 image 4, the extended risk table indicates that by the end of follow-up there was little difference in the number of participants experiencing an infection or infestations disorder between treatment groups. However, the event curves show that 50% of the placebo group experienced this event within approximately 100 days of randomisation, but it took until 160 days post randomisation for 50% of the mepolizumab group to experience the event.

Recommendations

The Kaplan–Meier plot with within-group confidence bands and extended risk table is recommended for specific events of interest to detect either a large between treatment group difference or a potential disproportionality over time, as ADRs are frequently time-dependent.

Potential amendments

For rare events, trialists may wish to reverse the vertical axis to display the cumulative proportion with the event to aid interpretation. It is also possible to create this plot in grayscale and use different line styles to differentiate between groups. Extensions to multiple events or multi-arm 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 to a common comparator or separate plots for different events.

Limitations

Kaplan–Meier plots only depict time-to-first event, failing to consider recurrent events. For clarity in presentation, they 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 (MCF) (see section: *5. Mean cumulative function*) is recommended. Some generic limitations of using time-to-event plots in this setting are provided in section: *Limitations applicable to time-to-event methods.*

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 is available here https://github.com/sarwarislam/kmunicate_stata.[23]

5. **Mean cumulative function plot – for recurrent events or a summary of the total burden of events**

Plot description

The mean cumulative function (MCF) 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 (figure 1 image 5, supplement 2 figure A.5). The 95% confidence interval bands show the precision of the within-group estimate. The risk table includes information on the number of participants that remain at risk of an event at discrete time points.

Implementation and interpretation

Over the first week post randomisation the mean number of events per participant is similar across treatment groups, but by day 20, a divergence becomes apparent (figure 1 image 5). In the paroxetine group, a mean of two events per participant were observed by day 20, but in the placebo group approximately 1.5 events per participant were observed by the same time-point. The plot of the MCF shows the participant burden of recurrent events, highlighting in this example that over follow-up paroxetine participants experience on average a greater number of events than placebo participants suggesting that there are some events associated to 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 '*x number(s) of an event*' will be experienced per participant by group. This can be provided as a summary to

demonstrate the burden of 'any event' as in the example presented here, or the recurrence of events of special interest. As highlighted in clinical feedback these plots are potentially very useful when investigating long-term therapies for chronic conditions and can provide valuable insight into periods the therapy might be considered 'safe' or 'well-tolerated'. When used to present data on 'any event' this plot serves as an alternative to the bar chart of counts that incorporates time. It may also serve as a useful summary of 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 per the Kaplan–Meier plot it is possible to create this plot in grayscale without loss of meaning. Extension to multi-arm studies or multiple events is potentially feasible but displaying multiple overlying confidence bands could make it incomprehensible, therefore in line with the recommendation for the Kaplan–Meier plot, trialists should consider only plotting the MCF (without confidence bands) and risk table or present separate plots for comparison of each intervention group to a common comparator or separate plots for different events.

Limitations

For clarity in presentation MCF plots are typically limited to one type of event at a time. More generic limitations and cautions of using time-to-event plots in the harm setting are provided in section: *Limitations applicable to time-to-event methods.*

Software

The MCF 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 MCF 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 section: *6. Survival-free ratio plot*). In addition, when using time-to-event methods for harm data, 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 et al. and include using appropriate estimates (e.g. Aalen-Johnson estimator or Fine and Gray method) to plot the cumulative incident function.[24]

6. **Survival ratio plot**

Plot description

This plot displays the ratio of non-parametric estimates of the survival probabilities (i.e. the probabilities for being event free in the harm setting) between treatment groups over time with a 95% confidence band. Unlike the Kaplan–Meier and MCF plots, this plot allows a direct comparison between treatment groups (figure 1 image 6, supplement 2 figure A.6). As the plot displays the ratio of survival probabilities over time, departures from unity indicate potential differences between treatment groups. The green horizontal bar at the bottom of the plot changes colour to red for the period that the confidence band excludes unity.[25]

Implementation and interpretation

Interpretation of the survival ratio plot (figure 1 image 6) depicts a point estimate indicating a greater risk of infection and infestation disorders in the placebo group compared to the intervention group with a value between 0.9 and 1 until day 40 dropping below 0.9 thereafter. Compared to the Kaplan–Meier plot, we can now see 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 (survival ratio = 1) 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 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. For events of specific interest where 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 absolute measure.**

Potential amendments

The example displays the ratio of survival probabilities estimated from the Kaplan–Meier method; alternatively, it could be used to display the difference in survival probabilities. Like both the Kaplan–Meier and MCF 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 only allows for time-to-first event, therefore it is not suitable for recurrent events. It is also limited to one type of event, however, in some situations it might be possible to add multiple estimates 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 it is important to consider competing risks when performing the underlying time-to-event analysis, further details of which are discussed in the section: *Limitations applicable to time-to-event methods*. 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 there is little evidence of 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 was confirmed in discussions with clinicians, who initially struggled interpreting this plot but indicated strong endorsement once 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 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 bands.

Recommendations for single continuous outcomes

7. **Line graph**

Plot description

In this plot, the markers display mean values and the 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 (figure 1 image 7, supplement 2 figure A.7). 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

In figure 1 image 7, there is an immediate drop in the mean eosinophil count after randomisation in the mepolizumab group and this 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 using 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 it may be better suited to depict clinical outcomes rather than blood markers where we are more often interested in the tails of the distribution (i.e. 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 e.g. when interest is in presenting descriptions of the distributions either means and standard deviations or medians and inter-quartile 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 not require use of colour.

Limitations

Changes in the tails of the distributions are usually of most interest when monitoring blood markers for harm and it may be difficult to see such changes using this plot. It is also unsuitable for skewed distributions so is better suited to present clinical outcomes rather than blood markers. Alternative plots for such data are presented below. Appropriate colour choices and line styles should be considered, particularly when adapting line graphs to multi-arm trials.

Software

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

8. **Violin plot**

Plot description

The hollow circle marker on the violin plot indicates the median value, the narrow rectangular boxes indicate the inter-quartile range and the lines extend from the box to the minimum and maximum points for each group at each time point. This is overlaid with kernel density plots, which summarise the distribution of the raw values (figure 1 image 8, supplement 2 figure A.8).

Implementation and interpretation

At time 0 (randomisation), the distributions were similar across treatment groups, but from week two onwards the distribution of the mepolizumab group values was narrower than the placebo group (figure 1 image 8). The distribution of the placebo group values remained 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 indicates a benefit for the mepolizumab group by reducing eosinophils.

Recommendation

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

Possible adaptations

In the current format, there is duplication of information in the mirrored kernel density plot. 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 multi-arm trials is not as space efficient as for the line graph. Kernel density estimates for some data may extend to values outside the plausible range e.g. figure 1 image 8, some kernel densities are below 0 for eosinophil counts, which are non-negative.

Software

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

9. **Kernel density plot**

Plot description

The kernel density plot displays the distribution of a continuous outcome. This can be at a single time point or for a derived change score e.g. the difference between the baseline value and maximum on treatment value (figure 1 image 9, supplement 2 figure A.9). Vertical reference lines can be included to indicate the upper and lower limits of normal values for the outcome.

Implementation and interpretation

Whilst figure 1 image 9 demonstrates that there is a similar distribution of values in the placebo and paroxetine groups that are within the normal range i.e. below 390 U/L, the plot

clearly demonstrates concerns for elevated alkaline phosphate values for some participants in the paroxetine group through the long right tail. This plot highlights the increased alkaline phosphatase levels in some participants taking paroxetine as an important event for closer monitoring in future trials or the post-marketing setting.

Recommendations

The kernel density plot is recommended to explore an outcome of interest at a specific time-point or a change score e.g. the change from baseline to a specific point in time or maximum change over the entire trial. The kernel density plot can be used to informally compare whole distributions between treatment groups and can highlight important differences in 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 it loses the information on repeated measures, only displaying information for one time point.

Software

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

Recommendations for multiple continuous outcomes

10. **Matrix of scatter plots**

Plot description

Multiple scatterplots of continuous outcomes. Each plot displays the relationship between values at two different time points, e.g. baseline values along the horizontal axis and the participant's maximum value over follow-up along the vertical axis (figure 1 image 10, supplement 2 figure A.10). The dashed lines represent the boundary between normal and abnormal thresholds

Implementation and interpretation

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). If there were more participants from the intervention group compared to control group in the top left quadrant this would be cause for concern. In figure 1 image 10 we can see there is a slightly higher number of individuals in the placebo arm (n=4) who had higher ALTs on treatment compared to baseline in contrast to the mepolizumab arm (n=2).

Recommendation

This plot is recommended in an exploratory setting to identify any outliers or patterns of interest. We suggest labelling outlying values with a participant identifier to assess if one or more participants have abnormal measurements across outcomes. This could be useful to monitor participants in ongoing studies and may also help raise signals for potential ADRs 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 colour schemes 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 impossible to distinguish individual points but transparency options could help with this.

Software

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

Areas for further development

Amongst the visualisations considered for displaying multiple time-to-event outcomes it was felt that the options available were poor. Whilst multiple Kaplan–Meier plots could be used to display information on a limited number of prespecified events of interest, there is still a gap in how to visualise multiple time-to-event outcomes simultaneously on the same plot. There were discussions about development of novel plots in this setting and this will be pursued in future work.

DISCUSSION

RCTs provide a valuable source of data to compare harm outcomes between treatment groups and can help identify potential signals for ADRs. However, there is evidence of suboptimal practices when reporting data on harm outcomes in clinical trial manuscripts. The CONSORT harms extension aimed to improve reporting and the recommendations from Lineberry et al. provided detailed examples to sit alongside CONSORT harms.[3 4] Both called for use of visualisations when reporting harm outcomes but did not give guidance on what visualisations would be helpful.

Principal findings

Researchers have called for guidance on appropriate methods for the analysis of harm outcomes and case studies detailing examples of use.[10] Our aim was to address this by providing 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, and clinicians. We have provided examples of the endorsed visualisations to communicate risks of harms in the RCT setting that can be used as an alternative to the widely used contingency tables. Our purpose in doing so is to increase the use of visualisations for harm outcomes in clinical trial manuscripts and reports, and ultimately 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, where 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 and the recommendations of Lineberry et al., as well as the more general guidance on the content of statistical analysis plans from Gamble et al.[3 4 26]

Ultimately, the choice of visualisation will depend on the outcome type (e.g., binary, count, time-to-event or continuous), the scenario (e.g., summarising multiple emerging events or one event of interest), the design of the trial (trials with more than two treatment groups require more care) and the purpose of the plot (e.g., to communicate information about the entire harm profile or convey a direct message about a particular event of interest). Therefore, it is for the statistician and clinical trial team to decide the most appropriate visualisation(s) for their data and objectives. It is likely that a combination of plots will be necessary, for example presenting both 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. The decision tree (figure 2) can be used by researchers to support their choices but this is not a one-size-fits-all approach, consideration is still required when deciding the most appropriate visualisations. It is also important to note that 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, and others include the standard error (SE), which is a measure of precision of the sample mean, or the 95% confidence interval, which is 1.96*SE, was also included. In these examples we have presented what was originally proposed, with context usually dictating which metric is the most suitable. The most suitable measure will be guided by the purpose of the plot.

Whilst 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, the colour scheme and symbols used, the axis scales and limits, text formatting, appropriate use of labels, and the number of groups being compared at once can all impact interpretation and understanding. Much has been written on these aspects, and we refer readers to the recent blog posts by Unwin and Rost, as well as lists of key principles for a good visualisation in the following references.[1 15 27-29]

Strengths and limitations

The predominance of statisticians over other researchers in the consensus group could be deemed a limitation of this work. However, statisticians are typically responsible for producing information on harms such as in tables or visualisations, and thus 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 their feedback sought 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 given they are likely to be the main consumers of such information, this allowed us to incorporate clarifications into the recommendations where necessary. Choosing clinicians who are active trialists will assist with dissemination and help us increase the likelihood of these plots being used in practice. Patients were not involved in this work as our focus was to identify the best plots to present in scientific journals with a predominant scientific readership. Our aim was to first provide guidance and tools to the authors of reports of RCTs. The next step that needs to be addressed is patient feedback. We did not consider use of interactive visualisations in these recommendations as we believe these fall into their own separate domain and require different considerations for appraisal (see Wang et al.[30]). Given the multifaceted, complex nature of data on harms and advances in the way we consume and access journal articles, interactivity could be highly advantageous for future projects.

Several novel plots were considered for endorsement in this work (for example the volcano and tendril plot shown in supplement 3 figures A.11 and A.15) but ultimately the appraisals revealed their inadequacies and a preference for more traditional plots. There was endorsement for two less commonly used plots, the survival ratio plot and the plot of the mean cumulative frequency (MCF), and we encourage use of such plots with clear explanations to ease interpretation. We particularly encourage use of the MCF 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 current lack of visualisations for presenting data on harm outcomes in the RCT setting, use of *any* visualisation of data on harms is arguably novel, especially for emerging events. Once the use of visualisations for harm outcomes is more common in the scientific press, this may increase the appetite for more innovative plots.

Whilst we suggest amendments to existing plots, the purpose of this work was not to develop new plots. However, it was clear there was a need for new approaches for some scenarios, particularly when interested in visualising multiple time-to-event outcomes or multiple continuous outcomes 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 avenue to explore and has previously been advocated by Chuang-Stein and Xia.[8] This would also serve as a readily available resource for dissemination. The CTSpedia Wiki page created by scientists from industry, academia and the Food and Drug Administration (FDA) goes some way towards this, serving as a repository of potential visualisations but provides limited direction on benefits of each plot, cautions of use and possible inferences to be drawn; it has also 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 these recommendations will improve reports of harm outcomes in clinical trial manuscripts, enabling clearer presentation of harm profiles and help identify potential signals for ADRs for further monitoring. Whilst we endorse each of the plots presented, we also highlight their limitations and provide examples of where their use would be inappropriate. We also caution users to practice care when creating and interpreting each plot. Though 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. 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 consumers of trial manuscripts can also appraise this information if they wish and is available to researchers wishing to undertake systematic reviews and meta-analyses of harms.[31]

Contributors

This project was conceived and designed by RP and VC. 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.

Acknowledgements

The authors would like to thank Dr. Will Stahl-Timmins for his inputs and guidance throughout this project and Rachel Evans (NWORTH CTU, Bangor University) for her contributions over the course of the consensus meetings.

Funding/Support

RP was funded by a NIHR Doctoral Research Fellowship to undertake this work (reference DRF-2017-10-131). SC is supported by an NIHR advanced fellowship (reference: NIHR300593). This paper presents independent research funded by the National Institute for Health Research (NIHR). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care. Infrastructure support for this research was provided by the Imperial College London NIHR Biomedical Research Centre (BRC). 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 the design, collection, analysis, interpretation, or writing of the manuscript.

The UK Clinical Research Collaboration Registered Clinical Trials Unit Statistics Operations Group supports and endorse this work.

Competing interest statement

All authors have completed the Unified Competing Interest form (available on request from the corresponding author) and declare: no support from any organisation for the submitted work other than funding described above, none of whom had a role in the design, collection, analysis, interpretation, or writing of the manuscript; 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.

Patient and public involvement

This work forms part of a wider research project that was developed with input from a range of patient representatives. There were no study participants directly involved in this work but the original proposal and patient and public involvement (PPI) strategy were reviewed by service user representatives (with experience as clinical trial participants and PPI advisors). We did not speak to patients directly for this piece of work as our focus was to identify the best plots to present in scientific journals with a predominant scientific readership. The next step that needs to be addressed is patient feedback.

Ethics

Ethics for this study analysis was granted from the Imperial Joint Research Compliance Office in February 2019. Ethics number 19IC5071.

Data sharing statement

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.

Transparency declaration

The lead author (the manuscript's guarantor) 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.

Figure titles and legends

Figure 1 title:

Figure 1: Endorsed visualisations

Figure 1 legend:
1. Dot plot; 2. Horizontal stacked bar chart of events by maximum severity; 3a. Bar chart of event counts for two treatment groups; 3b. Bar chart of event counts when > 2 treatment groups; 4. Kaplan-Meier plot; 5. Mean cumulative function plot; 6. Survival ratio plot; 7. Line graph; 8. Violin plot; 9. Kernel density plot; 10. Scatterplot matrix

Figure 2 title:

Figure 2: Decision tree to support selection of plot(s) to visualise data on harm outcomes

Figure 2 legend:
* Harm profile: a summary of all harm outcomes collected. Individual events: includes individual emerging events (including AEs and laboratory or vital sign data indicative of harm) and prespecified events of interest.

**Events of interest can include a single adverse event e.g. a headache or a single category of events that have been grouped together e.g. neurological body-system or an aggregated summary e.g. number of serious adverse events (SAEs)

Table 1: Key terms and definitions relating to harm outcomes

Term	Definition
Adverse (drug) reaction	Harm outcomes where a causal relationship between the intervention and event is "*at least a reasonable possibility*".[13 14]
Adverse event	Subset of harm outcomes that includes "*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*".[13]
Emerging	Non-prespecified events that are reported and collected during the trial and may be unexpected. Includes adverse events, and laboratory and vital sign data indicative of harm.
Harm outcomes	Encompasses individual emerging events and prespecified events of interest.
Harm profile	The summary or burden of the cumulative effect of all harm outcomes.
Prespecified	Individual events that are listed in advance as harm outcomes of interest to follow. They may be known or suspected to be associated to the intervention, or followed for reasons of interest.
Signal	Information that raises the possibility of a causal relationship between the drug and event.

Table 2: Endorsed plots and recommendations for use

Visualisations for summarising the entire harm profile (viewing differing multiple AEs)		
Outcome type	**Plot**	**Recommendation**
Binary	Dot plot	Use to present a comprehensive summary of the occurrence of multiple binary events
	Stacked bar chart	Use to present information on the occurrence and severity for multiple binary events
Count	Bar chart	Use to present information on event counts
Continuous	Matrix scatterplot	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 event(s) of interest* (viewing a single AE)		
Outcome type	**Plot**	**Recommendation**
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
	Survival ratio plot	Use as a signal detection tool to detect departures from unity to help detect potential signals for ADRs and alongside the Kaplan-Meier plot to incorporate a direct estimate of the between group difference for time-to-event outcomes
	Mean cumulative function plot	Use to display time-to-event information for recurrent events. Provides a visual summary of the time to expect '*x 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
	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
	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 may be a single adverse event e.g. a headache or a single category of events that have been grouped together e.g. neurological body-system or an aggregated summary e.g. number of serious adverse events (SAEs)

Table 3: Summary of characteristics to guide researchers in their choice of plot to visualise data on harm outcomes

Outcome type	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 and/or interest in exploring the distribution	Violin Plot
	Single outcome, at a single time-point	Kernel Density Plot
Time-to-event	Multiple outcomes	No suitable plot
	Single outcome	Kaplan–Meier Plot & Survival Ratio Plot
	Single, recurrent outcome	Mean Cumulative Function Plot

REFERENCE

1. Duke SP, Bancken F, Crowe B, et al. Seeing is believing: good graphic design principles for medical research. *Statistics in Medicine* 2015;34(22):3040-59. doi: doi:10.1002/sim.6549
2. Tufte ER. The Visual Display of Quantitative Information. 2nd ed. Chesapeake, CT: Graphics Press 2001.
3. Ioannidis JA, Evans SW, Gøtzsche PC, et al. Better reporting of harms in randomized trials: An extension of the consort statement. *Annals of Internal Medicine* 2004;141(10):781-88. doi: 10.7326/0003-4819-141-10-200411160-00009
4. 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 (Online)* 2016;355 (no pagination)(linn033422)
5. 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. *Clinical Trials* 2009;6(5):430-40. doi: 10.1177/1740774509344101
6. The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). ICH Topic E9 Statistical Principles for Clinical Trials, 1998.
7. Amit O, Heiberger RM, Lane PW. Graphical approaches to the analysis of safety data from clinical trials. *Pharmaceutical Statistics* 2008;7(1):20-35.
8. 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
9. Phillips R, Hazell L, Sauzet O, et al. Analysis and reporting of adverse events in randomised controlled trials: a review. *BMJ Open* 2019;9(2):e024537. doi: 10.1136/bmjopen-2018-024537 [published Online First: 2019/03/04]
10. 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(6):e036875. doi: 10.1136/bmjopen-2020-036875
11. Colopy MW, Gordon R, Ahmad F, et al. Statistical Practices of Safety Monitoring: An Industry Survey. *Therapeutic Innovation & Regulatory Science* 2019;53(3):293-300. doi: 10.1177/2168479018779973
12. Gelman A, Pasarica C, Dodhia R. Let's Practice What We Preach. *The American Statistician* 2002;56(2):121-30. doi: 10.1198/000313002317572790
13. Edwards IR, Biriell C. Harmonisation in pharmacovigilance. *Drug safety* 1994;10(2):93-102. doi: 10.2165/00002018-199410020-00001
14. The 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.
15. Unwin A. Why is Data Visualization Important? What is Important in Data Visualization? *Harvard Data Science Review* 2020;2(1) 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 [published Online First: 2012/08/22]
17. Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Statistics in Medicine* 2004;23(9):1351-75. doi: http://dx.doi.org/10.1002/sim.1761
18. Bradburn MJ, Deeks JJ, Berlin JA, et al. Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. *Statistics in Medicine* 2007;26(1):53-77.

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, et al. Longitudinal adverse event assessment in oncology clinical trials: The Toxicity over Time (ToxT) analysis of Alliance trials NCCTG N9741 and 979254. *The Lancet Oncology* 2016;17(5):663-70. doi: http://dx.doi.org/10.1016/S1470-2045%2816%2900038-3
21. Proctor T, Schumacher M. Analysing adverse events by time-to-event models: The CLEOPATRA study. *Pharmaceutical Statistics* 2016
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. *European journal of heart failure* 2014;16(1):33-40. doi: 10.1002/ejhf.29 [published Online First: 2013/12/18]
23. Morris TP, Jarvis CI, Cragg W, et al. Proposals on Kaplan–Meier plots in medical research and a survey of stakeholder views: KMunicate. *BMJ Open* 2019;9(9):e030215. doi: 10.1136/bmjopen-2019-030215
24. Proctor T, Schumacher M. Analysing adverse events by time-to-event models: the CLEOPATRA study. *Pharmaceutical Statistics* 2016;15(4):306-14. doi: 10.1002/pst.1758
25. Newell J, Kay JW, Aitchison TC. Survival ratio plots with permutation envelopes in survival data problems. *Computers in Biology and Medicine* 2006;36(5):526-41. doi: https://doi.org/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(23):2337-43. doi: 10.1001/jama.2017.18556
27. Rost LC. Thoughts & How To's How to pick more beautiful colors for your data visualizations. Chartable, 2020.
28. Vandemeulebroecke M, Baillie M, Margolskee A, et al. Effective Visual Communication for the Quantitative Scientist. *CPT: Pharmacometrics & Systems Pharmacology* 2019;8(10):705-19. doi: https://doi.org/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. *European Urology* 2020;78(1):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. *Statistics in Biopharmaceutical Research* 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(1):1028. doi: 10.1186/s13063-020-04903-0