
Review

Comparing the Value of Data Visualization Methods for Communicating Harms in Clinical Trials

Riaz Qureshi*, Xiwei Chen, Carsten Goerg, Evan Mayo-Wilson, Stephanie Dickinson, Lilian Golzarri-Arroyo, Hwanhee Hong, Rachel Phillips, Victoria Cornelius, Mara McAdams DeMarco, Eliseo Guallar, and Tianjing Li

* Correspondence to Dr. Riaz Qureshi, Department of Ophthalmology, School of Medicine, University of Colorado Anschutz Medical Campus, 1675 Aurora Court, Aurora, CO 80045 (e-mail: riaz.qureshi@cuanschutz.edu)

Accepted for publication August 17, 2022.

In clinical trials, harms (i.e., adverse events) are often reported by simply counting the number of people who experienced each event. Reporting only frequencies ignores other dimensions of the data that are important for stakeholders, including severity, seriousness, rate (recurrence), timing, and groups of related harms. Additionally, application of selection criteria to harms prevents most from being reported. Visualization of data could improve communication of multidimensional data. We replicated and compared the characteristics of 6 different approaches for visualizing harms: dot plot, stacked bar chart, volcano plot, heat map, treemap, and tendril plot. We considered binary events using individual participant data from a randomized trial of gabapentin for neuropathic pain. We assessed their value using a heuristic approach and a group of content experts. We produced all figures using R and share the open-source code on GitHub. Most original visualizations propose presenting individual harms (e.g., dizziness, somnolence) alone or alongside higher level (e.g., by body systems) summaries of harms, although they could be applied at either level. Visualizations can present different dimensions of all harms observed in trials. Except for the tendril plot, all other plots do not require individual participant data. The dot plot and volcano plot are favored as visualization approaches to present an overall summary of harms data. Our value assessment found the dot plot and volcano plot were favored by content experts. Using visualizations to report harms could improve communication. Trialists can use our provided code to easily implement these approaches.

adverse events; controlled clinical trials; data visualization; drug-related side effects and adverse reactions; harms; health communication; randomized controlled trials

Abbreviations: COSTART, Coding Symbols for a Thesaurus of Adverse Reaction Terms; ICE-T, Insight Confidence Essence Timing; IPD, individual participant data; MedDRA, Medical Dictionary for Regulatory Activities; MUDS, Multiple Data Sources in Systematic Reviews.

INTRODUCTION

Incomplete understanding of the potential harms of interventions threaten public health (1–6). “Harms” is an umbrella term endorsed by the Consolidated Standards of Reporting Trials Statement for unwanted effects of interventions, which is preferred rather than euphemisms such as adverse events, effects, reactions, risks, or complications (7). Harms are often summarized using frequencies and are presented in lengthy, often incomprehensible, tables.

Frequencies ignore the many other dimensions of the data important for patients, such as the timing of harms, duration of harms, whether harms may reoccur, and the severity and seriousness of harms (8). Moreover, reporting the occurrence of individual events (e.g., dizziness) neglects to inform stakeholders about important effects on categories or clusters of harms that can be observed when organized for analysis, such as by body systems as in Medical Dictionary for Regulatory Activities (MedDRA), a hierarchical standardized system for classifying harms, which replaced

Coding Symbols for a Thesaurus of Adverse Reaction Terms (COSTART) in 1999 (9, 10).

Trialists and journal editors should consider using visualizations rather than tables to present a more complete picture of the harms observed in trials. Patients and clinicians want more information than just the risk of common harms, and journal articles rarely provide the complete picture for evidence users (11–22). Patients are as interested in all potential harms as they are in potential benefits, and patients find it difficult to choose a subset of harms that would match those described in trial reports, picked in accordance with some form of selection criteria (13, 21, 23).

Selection criteria are the rules that clinical trialists and authors use to decide which harms will be presented in any given report (24–26). In previous studies, researchers have compared the individual participant data (IPD) from internal company documents for trials of gabapentin with the data presented in publications and found that although hundreds of unique harms can be collected in a single trial, less than a quarter of harms collected were ever reported in the subsequent trial publications (25–28). Other research has confirmed that this also occurs with serious harms (24–26, 29). Although manufacturers provide complete trial data to regulatory agencies, public reports about drugs—from publications to Food and Drug Administration approval packages and drug labels—do not enumerate all harms recorded in trials. For example, ClinicalTrials.gov includes tables of harms data in the reported results; however, investigators are only required to post those harms occurring in more than 5% of participants in trials of regulated products. Reporting each of the hundreds of harms for each treatment group might be overwhelming for patients and clinicians (12, 15, 16, 18, 22). Typically, authors use selection criteria based on numerical considerations, rather than importance, to choose a subset of harms that are most common. Selection criteria are highly variable within and between trial reports (24, 27). Because they are applied on the basis of data collected and analyzed, selection criteria introduce bias, and they dramatically affect the public's ability to identify and synthesize harms for interventions (27, 28, 30–34).

In an individual trial report, visualization of data may best meet the needs of patients and clinicians and help to overcome the limitation of selection criteria, which will improve public understanding of harms (35, 36). Tabulated numbers facilitate regulatory decisions, meta-analyses, and clinical guidelines, but the tables that appear in most journal articles are of limited value for these purposes; on the other hand, large tables that would be useful for those purposes are not effective for knowledge dissemination (35, 37–39). Neither tables reporting the frequency of hundreds of potential harms nor tables reporting just a few harms would capture what patients and clinicians need to know when they select an intervention (40, 41). In addition, to make these data more comprehensible, visualizations are useful when presenting multidimensional data, because most current approaches apply less restrictive selection criteria than tables, or they collapse potentially related events into higher-order categories of harms (e.g., MedDRA mid-level body systems) (36, 37, 42, 43).

In this study, we assessed the value of 6 visualizations of harms (41–43). Value describes a visualization's ability to communicate a complete picture and holistic understanding of the data (44–46). Value, in this context, goes beyond utility and can be assessed using 4 components: the time savings a visualization provides, the insights and insightful questions it spurs, the overall essence of the data it conveys, and the confidence about the data it inspires (44–46).

We generated 6 types of visualizations using patient-level data from a randomized controlled trial of gabapentin for neuropathic pain. A group of experts comprising the study authors and faculty and students of the Johns Hopkins Center for Clinical Trials and Evidence Synthesis compared the characteristics and value of these different approaches. Our objective was to review and assess these approaches from the perspective of content experts and to provide a reference toolbox for trialists to encourage easier and wider implementation of these approaches.

METHODS

Visualizations

From a recent scoping review (47) we selected 5 unique visualizations—stacked bar chart (48), dot plot (49), volcano plot (50), heat map (51), and tendril plot (52)—for presenting harms data. We also identified 1 additional visualization as a subtype of the heat map for evaluation—treemap—which has been used to present hierarchical medical data and summaries of harms (53–55).

In theory, any visualizations could present data on harms at any level; a visualization we produced with preferred terms (i.e., the lowest standardized terms for specific harms) could be used to present data on aggregated harms at a higher-order term. However, we have chosen to recreate these visualizations as presented in their source material. Consequently, our descriptions of the visualizations and methods to produce them, outlined in the following list, reflect the original versions.

- **Bar chart:** A bar chart in this context is a standard bar graph that depicts the relative occurrence of harms experienced across higher-order classifications of harms (e.g., mid-level or body-level systems). Specifically, the bars present the frequency of participants experiencing a harm to provide direct comparison of observed occurrence in a study. For additional information, the bars are broken down into different colors to present the occurrence by severity classification (i.e., mild, moderate, severe) to replicate the source material (49). This additional information, added by incorporating severity, technically makes this a stacked bar chart; however, in keeping with the source material, we have chosen to refer to this as a bar chart throughout this article.
- **Dot plot:** A dot plot is a 2-panel display of harms at the preferred-term level. The primary purpose of this visualization is to highlight potential signals by providing an estimate of the treatment effect and its precision. The first panel presents the incidence of each unique harm by treatment group, and the second panel presents

a summary measure of effect for each event and its corresponding 95% confidence interval. We created our dot plot with the relative risk as the summary measure to replicate the source material (50).

- **Volcano plot:** A volcano plot is a type of bubble plot that summarizes several characteristics of harms at the preferred-term level. Standard data elements represented in a volcano plot include the total frequency of each harm experienced in the trial (i.e., the total bubble area), the treatment arm with which there is a greater association (indicated by color and at the side of the figure), the statistical significance of the association (indicated by color saturation and position on the vertical axis), and the magnitude of effect (indicated by position on the horizontal axis). We created our volcano plot using the risk difference as the summary measure to replicate the source material (51).
- **Heat map:** A heat map presents data about the expected standardized effect for harms overall and across several selected subsets or subgroups of harms. The subgroups can be any that are available, but they should be distinct such that each “column” in the heat map shows a different set of harms that may be of interest to evidence users. A standardized effect is important in heat maps, because the risks can be variable across subgroups because of differences in denominators within subgroups, but the degree of uncertainty is difficult to show in a 2-dimensional field, thus the standardization aims to temper the lack of presentation of precision regarding estimates. The method of organization selected for the harms in a heat map is also important because it will change which inferences can be made more easily by looking at the figure. For example, when harms are arranged according to higher-order body systems, readers can see which specific events and body systems are most likely affected by the intervention. We created our heat map using a standardized risk difference, represented by color, and arranged our preferred terms by mid-level classifications, organized from most to least events in gabapentin, to replicate the source material (52).
- **Treemap:** A treemap is a subtype of heat map that presents data about the expected standardized effect for harms at the level of the preferred term in boxes, organized by their mid-level classification. As with the heat map, we used color to represent the standardized effect; however, unlike the heat map, the size of the box for each event also presents a dimension of data, which we chose to be the absolute count for events occurring in the intervention arm. The size of the preferred-term boxes also affects the size of the corresponding mid-level classifications, indicating which body systems are more commonly involved when gabapentin is used (55–57).
- **Tendrill plot:** A tendrill plot is a method of visually summarizing the timing, directionality, and magnitude of associations for harms. Each “tendrill” represents a preferred term for a harm, with the coloring of each point indicating false-discovery rate–corrected P values (Pearson’s χ^2 test for the hypothesis that the treatment

arms have the same proportion of events up to that event) and the size of each point being proportional to the total number of events for that preferred term. The path followed by the tendrill contains the information pertaining to event timing and direction of association (i.e., intervention or comparator). The time since randomization runs along each branch with the magnitude (or length) between points being proportional to the timing between each event. The center of the figure represents the start of the study and all tendrills begin moving directly upward, with each event shifting the direction of the tendrill—clockwise for events in the placebo arm or counterclockwise for events in the active arm—by some degree. The degree can be configured for optimal presentation and does not need to be the same for both arms: setting the degrees as proportional to the number of participants in each arm may be preferable with unbalanced treatment allocation (53).

Details about the production of each visualization, including how the data should be set up and the code to produce each figure, are provided in the supplemental R Markdown files (Web Appendices 1–5) (available at <https://doi.org/10.1093/aje/mxac005>) and a public GitHub repository (<https://github.com/rquresh/HarmsVisualization>). Table 1 presents the algorithms that underly the creation of each visualization.

Comparing visualizations

We created all visualizations in R (version 4.0.4, 2020; R Core Team, R Foundation for Statistical Computing, Vienna, Austria) using the same color palette and graphic options (e.g., font) so that all visualizations would be compared only on their content and not on our choices of elements to represent the data.

Comparing visualization characteristics. We compared the requirements for visualization generation according to how the data need to be set up, such as formatting and arrangement of the data set and the level of data (i.e., individual participant or aggregate). We extracted the characteristics of the visualizations according to how they organize the harms data (i.e., preferred term or higher-order terms) and the dimensions of the harms data they present. Dimensions of harms data include absolute occurrence (i.e., count), effect (e.g., risk) by treatment group, measure of effect (e.g., relative risk, risk difference), confidence intervals for estimates, significance test, severity of harms, and timing of harms. We recorded which dimensions are present in a visualization and which visualization elements are used in their presentation. We considered visualization elements to be the attributes of the data points in the figure, such as size, shape, position, color hue, and color saturation. If a dimension was not clearly presented through 1 of these attributes, we considered and discussed whether it was implicitly present (i.e., incorporated into an element presenting a different dimension), whether it could be included through a simple change (i.e., label or through 1 of the visualization elements), whether including it would necessitate a complicated change (i.e., restructuring the

Table 1. Steps to Produce Each Visualization

Visualization	Steps to Produce Figure
Bar chart	<p>A. For all (or a selection of) harms occurring over the trial and classify them according to a higher-order term of choice (e.g., either mid-level or body-level system)</p> <p>B. Code all harms by severity (as determined by trial investigators).</p> <p>C. Count the number of times each event and severity combination occurs for each arm in the trial.</p> <p>D. Generate the bar chart on the basis of the counts for all harms, grouped according to their higher-order term and colored according to severity, paneling the figure by the arm to compare between intervention and comparator.</p>
Dot plot	<p>A. For all (or a selection of) preferred terms for harms occurring in the trial and to calculate the incidence in each arm, followed by the desired summary measure of effect</p> <p>B. Sort the harms by the magnitude of effect.</p> <p>C. For all harms (<i>y</i>-axis), plot the incidence in each arm (<i>x</i>-axis) as 1 figure.</p> <p>D. For all harms (<i>y</i>-axis), plot the measure of effect and 95% confidence interval as a second figure.</p> <p>E. Join the figures into the resulting dot plot.</p>
Volcano plot	<p>A. For all (or a selection of) preferred terms for harms occurring in the trial and to calculate the desired measure of effect and corresponding <i>P</i> value (unadjusted or adjusted for multiple testing using a method like false discovery rate)</p> <p>B. Compute $-\log_{10}(\text{raw } P \text{ value})$ for the <i>y</i>-axis.</p> <p>C. Plot the harms with the $-\log_{10}(P \text{ value})$ as the <i>y</i>-axis (i.e., higher on the figure corresponds to greater statistical significance) and the measure of effect as the <i>x</i>-axis (i.e., estimates to the right of null are more common in the intervention arm and estimates to the left are more common in the control arm).</p> <p>D. Size the bubbles according to total events.</p> <p>E. Color and saturate according to the magnitude and significance of treatment effect.</p> <p>F. Add a reference line for statistical significance (i.e., $-\log_{10}(0.05)$ data for unadjusted for multiple testing and $-\log_{10}(\alpha)$ for data adjusted for multiple testing)</p>
Heat map	<p>A. For all (or a selection of) preferred terms for harms occurring in the trial and to compute the standardized difference for all events "overall" with the following formula: $(p_t - p_c) / (\sqrt{p_t \times (1 - p_t) / n_t + p_c \times (1 - p_c) / n_c})$</p> <p>B. Select subgroups of interest or by availability and keep only the events meeting those criteria, then calculate the standardized difference for events in each of these subgroups. Subgroup examples include:</p> <ol style="list-style-type: none"> i. Sex: female, male ii. Seriousness: serious harms, nonserious harms iii. Severity: moderate and severe harms, severe harms iv. Recurrence: single-episode harms, multiple-episode harms v. Relatedness to intervention: likely related, possibly related, definitely related <p>C. Plot the specific harms on the <i>y</i>-axis with the different subgroups on the <i>x</i>-axis.</p>
Treemap	<p>A. For all (or a selection of) preferred terms for harms occurring in the trial and to compute the standardized difference for all events "overall" with the following formula: $(p_t - p_c) / (\sqrt{p_t \times (1 - p_t) / n_t + p_c \times (1 - p_c) / n_c})$</p> <p>B. Plot the specific harms grouped according to their corresponding mid-level terms using the Treemap function.</p> <p>C. Size the boxes according to number of events in the intervention arm.</p>
Tendrill plot	<p>A. For all (or a selection of) preferred terms for harms occurring at least 2 times in the trial (i.e., the minimum to create a single vector, or whatever the desired threshold of occurrence may be): sort all events by time (i.e., days since randomization).</p> <p>B. Calculate the length for each vector as the time between subsequent events. Vector length = 0 for an event reported on the same day as the previous event.</p> <p>C. Calculate the angle for each vector. The angle is the cumulative sum of all angles up to that event: the angle rotates clockwise for events on the placebo arm and counterclockwise for events on the active arm. Vectors of zero length still contribute to angular changes.</p> <p>D. Add the vectors together cumulatively for each harm (i.e., each subsequent vector begins where the previous vector ends), starting from $X = 0, Y = 0$. The resulting sequence of vectors creates the tendrill for that harm.</p>

Abbreviations: n_c = number of participants in control arm; n_t = number of participants in treated arm; p_c = proportion with event in control arm; p_t = proportion with event in treated arm.

visualization), or whether it would be impossible to include in the visualization given the current visualization design.

Comparing visualization value. We compared the value of visualizations using a modified ICE-T framework, a methodology that uses a heuristic approach to evaluating the value of multiple visualizations (45). The ICE-T methodology is meant to complement other methods for evaluating usability and value by providing a high-level, value-driven assessment (45, 46). The ICE-T framework comprises multiple statements that are assessed on a Likert scale of agreement and fall under 4 components: insight, time, essence, and confidence. Respectively, these components capture whether each visualization 1) supports insights, 2) supports faster and more efficient understanding of the data, 3) communicates the overall essence of the data set, and 4) gives the user confidence in their understanding of the data (45, 46). The original ICE-T approach was designed to be used by visualization experts, who serve as evaluators. We modified the original approach by removing heuristics that did not apply to our visualizations (e.g., we excluded heuristics for assessing interactive features, because our visualizations are static representations), and we rephrased the wording of some of the heuristics to make them more applicable and accessible to evaluators who had clinical trials expertise. Web Appendix 6 is our revised ICE-T survey form.

We distributed the modified ICE-T survey form to all members of our author team ($n = 12$) and to faculty and students of the Center for Clinical Trials and Evidence Synthesis at Johns Hopkins University who attended a presentation on the research ($n = 20$). Our author team comprises experts in the field brought together by common research interests in the analysis and reporting of harms in trials. We collected responses anonymously and thus were unable to separate responses from authors and nonauthors. All invited respondents are content experts and include clinical trialists, trial biostatisticians, biomedical data visualization, and pharmaco-epidemiologists. We asked all respondents to complete the modified ICE-T survey for each of the 6 visualizations. We ordered the presentation of visualizations on the basis of when the source materials were chronologically published. Each visualization was accompanied by a background paragraph describing its components. This research was determined to be nonhuman subjects research by the Johns Hopkins Bloomberg School of Public Health Institutional Review Board.

We aggregated scores using the method described by Wall et al. (44)—averaging the scores for individual heuristics by domain for each respondent—and we tabulated the results to show the overall strength of each visualization with respect to each of the 4 components (45). The ICE-T survey uses a 7-point Likert scale, and a higher score reflects greater value (45).

Data

The data used in this methodologic study are the IPD from a randomized controlled trial randomly selected from 6 randomized controlled trials of gabapentin obtained as part of a review from the Multiple Data Sources in System-

atic Reviews (MUDS) Study (completed in 2017) (23, 25, 26, 56–59). MUDS was a methodological project funded by the Patient-Centered Outcomes Research Institute, and members of our team (E.M.W. and T.L.) were investigators on the MUDS study. Briefly, in 2008, internal company documents including IPD from 6 trials, email correspondence, memoranda, study protocols, and reports were made public after litigation was brought and won against Pfizer for the intentional misreporting of randomized controlled trials of gabapentin to promote off-label use (56–58). Data from these trials have primarily been explored with regard to efficacy outcomes. The research on harms from these trials, however, has only ever been reanalyzed with regard to those which were ever or never reported in publications (25, 26).

We have analyzed these data in their entirety in fulfillment of a grant from the Restoring Invisible and Abandoned Trials initiative. These harms were coded by the original trial investigators in a hierarchy—from common terminology to standardized preferred terms and associated higher-order terms—using COSTART. Details of our methods for analyses of these trials can be found in the published protocol (60).

RESULTS

Web Figure 1 presents the visualizations produced and compared for trial 945–210.

Evaluation—modified ICE-T

From the 32 invited experts, we collected a total of 17 responses to our modified ICE-T survey, exploring the value of the 6 visualizations. Web Figure 2 presents the mean scores for each domain of each visualization by respondent, organized in descending overall average score, with 4 being the middle and higher scores indicating greater value and lower scores indicating less value. In general, respondents rated the dot plot and volcano plot highly overall, with particularly high scores for both time and ability to convey the essence of the data. The treemap also had an overall score above 4 and scored high on insights about the data but scored low in terms of the confidence in the underlying data. The bar chart and heat map were rated average overall, although the bar chart had more scores around 4, whereas the heat map scored high for insight but low for time and confidence. The tendril plot was rated the lowest of all visualizations across all domains, suggesting relatively poor comprehensibility, even among experts.

The last question of our survey asked respondents to rank the visualizations from most preferred to least preferred approach for presenting summary data on harms. Figure 1 and Table 2 present the rankings of each visualization, ordered from most preferred to least preferred by mean rank. In Figure 1, dark blue and dark red correspond with a ranking of 1 and 6, respectively, with 1 being the most preferred. The results were consistent with the ICE-T scores, with the dot plot and volcano plot ranked most preferred and the tendril plot ranked least preferred type of visualization. Overall, participants did not indicate strong preferences among the treemap, bar chart, and heat map.

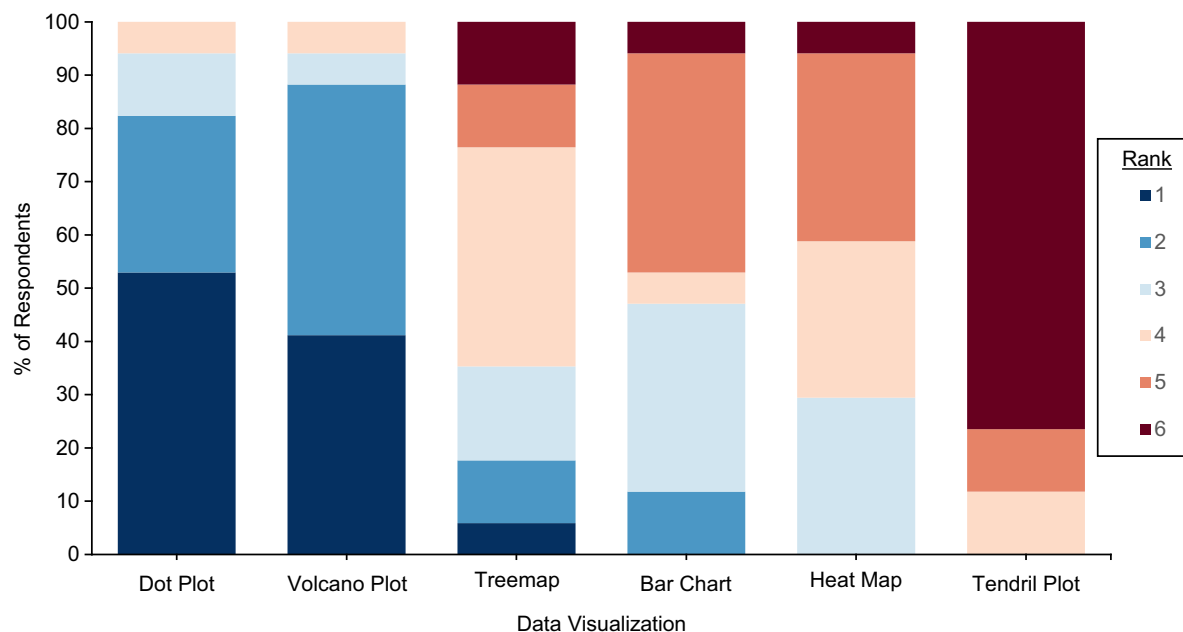


Figure 1. Proportion of survey respondents ($n = 17$) ranking each visualization in order of preference for presenting summary harms data collected in a trial. Lower numbers correspond to a higher rank and greater preference, with 1 being the most preferred.

Comments on visualizations

As part of our value assessment, we included a free-text question to capture any comments respondents had about each visualization. Comments about the dot plot ($n = 6$) focused on the readability of the visualization, given the number of harms and amount of data fit into the space. One responder wrote, “There are a lot of values and it’s difficult to interpret so many effects.” The comments on the bar chart ($n = 6$) suggested that a way of representing the difference between the groups such as through better bar arrangement would be helpful to better understand the comparison. Comments on the volcano plot ($n = 5$) did not have any common themes, but it was noted that there was duplication of information, with multiple visualization elements representing the same dimension, and that harms

with the same set of data overlapped and “disappeared” from the visualization (i.e., 2 harms with the same number of events in both groups would occupy the same space on the figure). The comments on the heat map ($n = 7$) noted a lack of clarity in the direction of treatment effect (e.g., “Was not clear what color favored which arm”) and the information conveyed, suggesting that there may be too much information presented in 1 visualization (e.g., “I find this very cluttered and challenging to read and interpret”). The comments on the treemap ($n = 2$) described it as a good holistic assessment of terms and groups of terms but lacking in precision and clarity in exact details. Last, the comments on the tendril plot ($n = 10$) focused on the need for labels for harms and the challenge in interpretation. One responder wrote, “I find this way too challenging. Maybe a way to look at a few focused questions, but a terrible starting point.”

Table 2. Mean and Median Ranking Scores of 6 Visualizations From Most (1) to Least (6) Preferred

Visualization	Mean (SD)	Median (IQR)
Dot plot	1.7 (1.2, 2.2)	1 (1, 2)
Volcano plot	1.8 (1.3, 2.2)	2 (1, 2)
Treemap	3.8 (3.1, 4.5)	4 (3, 4)
Bar chart	3.9 (3.3, 4.6)	4 (3, 5)
Heat map	4.2 (3.7, 4.7)	4 (3, 5)
Tendril plot	5.6 (5.3, 6.0)	6 (6, 6)

Abbreviations: IQR, interquartile range; SD, standard deviation.

Generation of data visualizations

To generate different visualizations, investigators need to set up data in different ways, which might be “wide” or “long” formatting and might be at the level of the individual participant or all participants assigned to the same group. All visualizations make use of IPD from the trials to ensure the completeness of the data (i.e., that all collected harms are available). However, the tendril plot is the only visualization that requires IPD as the input data for generation; all the other visualizations are created using aggregated data. Aggregated data are the summary counts, risks, and estimates of effect for every captured harm.

Once the corresponding data have been aggregated for all the harms in the trial, it is easy to select the appropriate

Visualization	Organization of Harms		Characteristics of Harms Data Visualized						
	Preferred Terms	Higher-Order Terms	Absolute Occurrence of Harm	Estimate by Treatment Group	Measure of Effect	Precision of Estimates	Significance Test	Severity	Timing
Dot plot	++		-	++	++	++		--	-
Bar chart	-	++	++		-	-	-	++	-
Volcano plot	++		++	--	++	-	++	--	-
Heat map	++	++	-	--	++	+		++	-
Treemap	++	++	-	--	++	+		--	-
Tendrill plot	++		++	--		-	++	--	++

Definition

++ Present and clear

+ Present but implied

| Could easily be added

- Difficult to add (requires change)

-- Not possible with graphic

Figure 2. Presence or absence of data characteristics in 6 visualizations of harms collected in trials. Characteristics of harms data may be clearly present in a graphic (blue ++), not possible to include with the current design (red --), or somewhere in between.

variables for the desired visualization. The supplemental R Markdown files provide an example data structure including all necessary columns to produce each figure (Web Appendices 1–5).

Generating each visualization is easy in R using existing packages, so we did not create any new packages. Four of the 6 visualizations—dot plot, bar chart, volcano plot, and heat map—can be produced using standard graphical commands for common plots within the *ggplot2* package (e.g., `geom_point`, `geom_tile`). The treemap and tendrill plot require the *Treemapify* and *Tendrill* packages, respectively (61, 62).

DISCUSSION

We compared the characteristics of 6 different visualizations and found that data visualization is a valuable way to present multiple dimensions of harms data simultaneously and to summarize the overall profile of harms for an intervention. Our assessment of value by content experts found the dot plot and volcano plot were valued the most highly of the visualizations we assessed, although there are several different approaches that can meet the needs of investigators and trialists. The dot plot and volcano plot scored very highly overall, although researchers should explore ways of improving the confidence about the data that they impart (i.e., trust about the data, its value, and context), because this was the lowest scoring domain. The treemap, bar chart, and heat map were rated more modestly by the respondents and could be worth exploring further by trialists as approaches to communicating overall summaries of harms by patients and clinicians. The tendrill plot was the least favored of all the visualizations. Respondents found it difficult to understand, although multiple respondents suggested it could be valuable for detailed exploration of specific harms.

Our study is complementary to previous research on visualizing harms. In particular, we assessed value from the perspectives of trialists and biostatisticians who are most familiar with trial data and who are the people typically responsible for producing these plots and communicating

interpretation to the trial team. This study thus serves as a reference and guide for investigators who want to know which visualization best suits their needs. We have supplied the R code to produce these visualizations so trialists can easily implement these approaches using their own data. Whereas early assessments and reviews of data visualization techniques focused largely on describing different ways of representing statistical aspects of the data or detailed assessments of a few harms of interest, more recent research has focused on comparing visualizations with tables as a method to present harms and developing consensus from trialists on the use of visualizations (40, 48, 49, 63–65). Our discussion of characteristics and the value assessment by content experts is a logical precursor to working with end users of the visualizations, such as patients and clinicians. This research will be used to support the development of consensus guidelines from trialists as well as the full-scale evaluation of value by other stakeholders.

Characteristics of data visualizations

Here, we discuss the characteristics or dimensions of harms data apparent in each visualization as we have reproduced them from their source material, and we discuss how we believe they could be modified to include other characteristics as presented in Figure 2. The colors and symbols in Figure 2 correspond with whether we considered the characteristic to be clearly present in the visualization (blue ++), not possible in the graphic without drastically changing the structure (red --) or something in between.

Organization of harms. Preferred terms and higher-order terms respectively refer to the lowest standardized term for a specific harm and the mid to upper levels of a hierarchical system such as COSTART or MedDRA into which the preferred terms can be categorized (e.g., organ or body system). Five of 6 visualizations propose presenting data for harms at the preferred-term level and 1 presents data for harms at the level of higher-order terms. The heat map and treemap present data on harms at the preferred-term level

but are organized according to higher-order terms to help identify patterns in affected body systems. This organization by higher-order terms could be incorporated into the dot plot, the volcano plot, or the tendril plot by grouping the preferred terms (*y*-axis), coloring the bubbles, or changing the symbol for each tendril, respectively, according to higher-order terms. For the bar chart, including data on a secondary level of harms would require additional visualization elements or changing the structure of the visualization.

Dimensions of harm data. The absolute occurrence (i.e., count) of harms can be seen in 5 of the 6 visualizations: the heights of the bars in the bar chart, the size of the points in both the volcano plot and tendril plot, a higher position on the *y*-axis of the heat map for the mid-level groupings, and the size of the boxes in the treemap. The dot plot could be modified to include the absolute occurrence of harms by the addition of a data table or a label with this information to each row on the *y*-axis. An estimate for the effect of harms (e.g., risk) by treatment is only presented in the dot plot, although the bar chart presents the count of harms data by treatment group, which could be modified to instead present an estimate of effect.

An estimate for the measure of effect (i.e., any between arm treatment effect such as a relative measure for intervention vs. control, including odds ratio, risk ratio, hazard ratio, and risk difference) is presented through visualization elements in the dot plot (position of points along the *x*-axis on the right side of the figure), the volcano plot (position of points along the *x*-axis), and the heat map and treemap (the color of each tile). The tendril plot could incorporate a measure of effect by coloring each tendril according to its estimate; however, this would replace the current use of color to represent the *P* value. Alternatively, the tendril plot could be modified to include a label for the overall measure of effect of each tendril, an approach that could also be used in the bar chart.

The precision of the relative effect estimate (i.e., any method of accounting for the precision in the estimate, whether using a confidence interval or through standardization of the effect size) between treatment groups is directly represented only in 1 visualization: the width of the bars around each estimate in the dot plot. The heat map and treemap present a standardized effect size (i.e., measure of effect divided by the standard error) and thus the precision is incorporated into the tile color, but precision is not represented with its own visualization element. The volcano plot and tendril plot could both potentially incorporate the precision of estimates using the visualization element of color intensity or shade of a point, but this would not be directly interpretable as in the dot plot; using color this way would replace the use of color to represent the *P* value. As with the estimated measures of effect, the bar chart presents data by treatment group, and we did not identify an element that could be modified or added to represent the precision of estimates without adding another feature.

The statistical significance of an estimate (specifically, the *P* value) is directly presented in both the volcano plot and the tendril plot. A label could be added to each row on the *y*-axis for both the dot plot and the heat map to indicate

the associated *P* value. Although the bar chart presents only counts for higher-order classifications of harms and does not present any measures of effect, it would also be possible to add an indicator to each column on the *x*-axis (e.g., a symbol or highlight of the axis text) for those where there is a significant effect.

The severity of harms is presented directly in the bar chart as the number of times that harms are mild, moderate, or severe and in the heat map as a subset of harms such that the standardized effects are shown for only harms reported as moderate or Severe. We did not identify a way that the other 4 visualizations could incorporate severity; harms can have varying degrees of severity and could be represented categorically, but other visualizations present each preferred term only once.

The timing of harms is incorporated as a direct element for only the tendril plot, represented by the length of the vectors between each event. Although none of the other visualizations include an element to represent timing, timing could be included by paneling the visualizations by time periods. That is, paneling would create a separate instance of a visualization for harms occurring within time periods of interest (e.g., weeks 1–4, weeks 5–8, weeks 9–12, and weeks 13–16). Paneling could create a series of visualizations that give insight into which harms are more frequent earlier or later after randomization.

Important considerations when creating visualizations

In assessing the characteristics of these visualizations and their value, we identified several important considerations and decisions that must be made by investigators. These aspects should also be considered by journal editors who influence how trialists present data. These include decisions regarding analyses and interpretation of harms results, target audience for visualizations, the inclusion of supporting materials, and whether to present all harms collected in a trial or a subset meeting some criteria.

A major challenge in analyzing harms arises because some trials are not designed to assess harms, and many harms occur infrequently in trials designed to assess potential benefits. Low statistical power and multiple hypothesis testing might lead to both false-positive and false-negative results in individual trials (8, 41, 47). Drug trials frequently focus on potential benefits, and journal articles typically provide only abbreviated data on harms; tables in journal articles are usually inadequate for determining whether harms are causally related to the drug's use. When conducting analyses, investigators need to consider whether to use some adjustment for multiple testing, such as a Bonferroni correction or false discovery rate (41, 66–68). Similarly, when creating a visualization to present harms data and the *P* value is 1 of the dimensions being presented, it is important to consider which values are presented, because visualizing the *P* value or confidence intervals may lead to overinterpretation of the results for harms (40, 41, 47, 50, 51, 65). Unless a harm is prespecified and systematically assessed in a trial, its analysis should be considered exploratory and should not be interpreted under standard hypothesis-testing assumptions (8, 40, 41, 50, 51, 68).

Investigators and editors should also consider the target audience for each visualization. Depending on whether the investigators are trying to communicate harms to patients, clinicians, or other stakeholders, different levels of health literacy and numeracy might be expected (40, 44, 69–71). To effectively communicate harms, a visualization must be understood by the target audience (36, 40, 44, 70, 71). For example, the dot plot and volcano plot scored high on the time and essence domains and were ranked as the most preferred visualizations in our value assessment by content experts. Compared with the tendril plot, which was ranked last and had low scores for all 4 domains, the dot plot and volcano plot are easier to understand and likely would better serve a nontechnical audience (40, 65).

Regardless of the target audience, investigators must also decide whether to include supplemental materials to accompany visualizations (40, 41, 65). Although visualizations are good for communicating an essence and overall summary of data, tables are better for presenting precise information (35, 37, 39, 40). If investigators choose to present the harms using visualizations in trial publications, they could also include supplemental tables with all the data informing the visualization so systematic reviewers and guideline developers can reuse the data (40, 41, 65).

An additional and important consideration for any visualization is whether to present data on all harms collected in the trial or to present a subset of harms and report all harms in supplementary tables. Patients and health care providers want to know more about harms than journal articles report; however, some visualizations may become overcrowded and difficult to read if there are too many harms or dimensions being presented (40, 65). Several comments from respondents in our study noted that visualizations were cluttered when they presented data on all harms. Ideally, as technology and visualization approaches evolve, interactive online visualizations might become common, allowing users to apply their own filter(s) so they have access to all data but are not overwhelmed and can focus on harms of personal interest (72, 73). The specifics of how to select harms to present is an area of research that requires more work. Harms could be selected for reporting by prespecifying those that might be expected (e.g., based on preclinical studies or studies of similar drugs) or harms of greatest importance to patients (e.g., serious harms, harms that led to discontinuation of treatment, harms identified in core harm-outcome sets developed by consensus) (74). Patient involvement is crucial in making these selections.

Future research directions

In addition to identifying considerations for creating visualizations, our evaluation of these visualizations revealed many opportunities for future research. There have been several reviews of different approaches to presenting harms data, but, to our knowledge, ours is the first that explicitly compares the characteristics of harms data and the respective value of these visualizations (40, 48, 49, 63–65). We included study authors in our sample for assessing the value of visualizations, because all are content experts and fit our target audience; however, further studies are needed to deter-

mine whether these results generalize to other researchers, and whether clinicians and patients have similar or different preferences.

Our study is an important first step toward increasing the use of visualizations in trial publications and improving the communication of harms to evidence users by providing a reference for trialists and those who will be creating visualizations. Consensus methods involving multiple rounds of surveys and interviews with clinicians and clinical trialists could be used to develop recommendations about how and when specific visualizations should be used in journal articles. Such a consensus would also provide guidance on which visualizations may be best used for identifying harms associated with administration of a drug versus adverse events occurring in the background. Efforts are underway to create a consensus on specific visualization uses and presentation of results in journal articles from these stakeholders (65). Additionally, although we did not include important targeted stakeholder groups such as patients and clinicians in our assessment, now that we better understand the characteristics of these visualizations and have a measure of their comparative value from experts, we will direct efforts toward evaluating the utility of these visualizations to meet their needs. Researchers could explore visualizations and other solutions for presenting relationships between harms (e.g., co-occurrence) and multiple arms (e.g., trials with more than 2 arms).

Additional evaluation of visualization approaches should also include a feedback loop to improve them, for example by asking stakeholders how they would change the visualizations. The ability to interact with a visualization and select harms that meet one's own criteria of interest may be considered valuable and worth incorporating in future visualizations (37, 38, 44, 72, 73). Similarly, if there is a dimension that has not been visualized but would be valuable to stakeholders, such as the duration of harms or which harms are likely to co-occur, it would be worth exploring how best to incorporate such information with stakeholders to ensure their needs are met. Another aspect that should be assessed in future evaluation is whether stakeholders can identify key harms that have been established as important by expert consensus.

CONCLUSIONS

Data visualization has the potential to reduce bias in the reporting of harms and to present a more holistic picture of harms, compared with current methods. Of the 6 approaches we evaluated for visualizing harms observed in clinical trials, we found a strong preference for 2 (i.e., dot plot and volcano plot) and dislike for 1 (i.e., tendril plot). Clinical trialists should use visualizations to present data on multiple dimensions for harms. Systematic reviewers could apply similar approaches to summarizing harms results from meta-analyses, depending on the data they have available from primary studies. Journals should consider making such visualizations standard recommended components for trials. Researchers could focus on the evidence user's perspective to ensure that visualizations used in practice will communicate harms appropriately and adequately to stakeholders.

ACKNOWLEDGMENTS

Author affiliations: Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, United States (Riaz Qureshi, Mara McAdams DeMarco, Eliseo Guallar); Department of Biostatistics, Indiana University School of Public Health, Bloomington, Indiana, United States (Xiwei Chen, Stephanie Dickinson, Lilian Golzarri-Aroyo); Department of Biostatistics, University of Colorado Anschutz Medical Campus, Denver, Colorado, United States (Carsten Goerg); Department of Epidemiology, Indiana University School of Public Health, Bloomington, Indiana, United States (Evan Mayo-Wilson); Department of Biostatistics and Bioinformatics, Duke University, Durham, North Carolina, United States (Hwanhee Hong); Faculty of Medicine, Imperial College London, London, United Kingdom (Rachel Phillips, Victoria Cornelius); and Department of Ophthalmology, University of Colorado Anschutz Medical Campus, Denver, Colorado, United States (Tianjing Li). Dr. Qureshi is now at the Department of Ophthalmology, University of Colorado Anschutz Medical Campus, Denver, Colorado, United States.

This work was supported by funds established for scholarly research on reporting biases at Johns Hopkins by Greene LLP (to R.Q.); Laura and John Arnold Foundation and the Restoring Invisible and Abandoned Trials Initiative (to X.C., S.D., L.G.-A., and E.M.-W.).

The data are available from the corresponding author.

Presented at the 2022 Society for Clinical Trials Annual Meeting, May 15–18, 2022, in San Diego, California.

Conflict of interest: none declared.

REFERENCES

- Ioannidis JP. Adverse events in randomized trials: neglected, restricted, distorted, and silenced. *Arch Intern Med.* 2009; 169(19):1737–1739.
- Pitrou I, Boutron I, Ahmad N, et al. Reporting of safety results in published reports of randomized controlled trials. *Arch Intern Med.* 2009;169(19):1756–1761.
- Dwan K, Gamble C, Williamson PR, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. *PLoS One.* 2013; 8(7):e66844.
- Golder S, Loke YK, Wright K, et al. Reporting of adverse events in published and unpublished studies of health care interventions: a systematic review. *PLoS Med.* 2016;13(9): e1002127.
- Loke YK, Derry S. Reporting of adverse drug reactions in randomised controlled trials - a systematic survey. *BMC Clin Pharmacol.* 2001;1:3.
- Landefeld CS, Steinman MA. The Neurontin legacy — marketing through misinformation and manipulation. *N Engl J Med.* 2009;360(2):103–106.
- Ioannidis JP, Evans SJ, Gøtzsche PC, et al. Improving patient care better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med.* 2004;141(10):781–788.
- Ma H, Ke C, Jiang Q, et al. Statistical considerations on the evaluation of imbalances of adverse events in randomized clinical trials. *Ther Innov Regul Sci.* 2015;49(6):957–965.
- US Food and Drug Administration. *National Adverse Drug Reaction Directory: “COSTART” (Coding Symbols for Thesaurus of Adverse Reaction Terms)*. Washington, DC: US Department of Health, Education, and Welfare; 1970.
- International Conference on Harmonization. MedDRA Medical Dictionary for Regulatory Activities. International Council for Harmonisation website. 2021. www.ich.org/page/meddra. Accessed May 5, 2021.
- Berry DC, Michas IC, de Rosis F. Evaluating explanations about drug prescriptions: effects of varying the nature of information about side effects and its relative position in explanations. *Psychol Health.* 1998;13(5):767–784.
- Büchter RB, Fechtelpeper D, Knelangen M, et al. Words or numbers? Communicating risk of adverse effects in written consumer health information: a systematic review and meta-analysis. *BMC Med Inform Decis Mak.* 2014; 14(1):76.
- Berry D, Michas I, Bersellini E. Communicating information about medication side effects: effects on satisfaction, perceived risk to health, and intention to comply. *Psychol Health.* 2002;17(3):247–267.
- Enlund H, Vainio K, Wallenius S, et al. Adverse drug effects and need for drug information. *Med Care.* 1991;29(6): 558–564.
- Lang S, Garrido MV, Heintze C. Patients’ views of adverse events in primary and ambulatory care: a systematic review to assess methods and the content of what patients consider to be adverse events. *BMC Fam Pract.* 2016;17(1):1–9.
- Entwistle VA, Mello MM, Brennan TA. Advising patients about patient safety: current initiatives risk shifting responsibility. *Jt Comm J Qual Patient Saf.* 2005;31(9): 483–494.
- Duclos CW, Eichler M, Taylor L, et al. Patient perspectives of patient-provider communication after adverse events. *Int J Qual Health Care.* 2005;17(6):479–486.
- Tarn DM, Wenger A, Good JS, et al. Do physicians communicate the adverse effects of medications that older patients want to hear? *Drugs Ther Perspect.* 2015;31(2): 68–76.
- Basch E. The missing patient voice in drug-safety reporting. *N Engl J Med.* 2010;362(10):865–869.
- Berry DC, Raynor DK, Knapp P, et al. Patients’ understanding of risk associated with medication use: impact of European Commission guidelines and other risk scales. *Drug Saf.* 2003;26(1):1–11.
- Mcneal TM, Colbert CY, Cable C, et al. Patients’ attention to and understanding of adverse drug reaction warnings. *Patient Intell.* 2010;2:59–68.
- Franklin L, Plaisant C, Shneiderman B. An information-centric framework for designing patient-centered medical decision aids and risk communication. *AMIA Annu Symp Proc.* 2013;1:456–465.
- Mayo-Wilson E, Fusco N, Hong H, et al. Opportunities for selective reporting of harms in randomized clinical trials: selection criteria for non-systematic adverse events. *Trials.* 2019;20(1):553.
- Mayo-Wilson E, Fusco N, Li T, et al. Harms are assessed inconsistently and reported inadequately part 2: non-systematic adverse events. *J Clin Epidemiol.* 2019;(113): 11–19.
- Mayo-Wilson E, Fusco N, Li T, et al. Harms are assessed inconsistently and reported inadequately part 1: systematic adverse events. *J Clin Epidemiol.* 2019;(113):20–27.

26. Mayo-Wilson E, Fusco N, Li T, et al. Multiple outcomes and analyses in clinical trials create challenges for interpretation and research synthesis. *J Clin Epidemiol.* 2017;(86):39–50.
27. Mayo-Wilson E, Li T, Fusco N, et al. Cherry-picking by trialists and meta-analysts can drive conclusions about intervention efficacy. *J Clin Epidemiol.* 2017;(91):95–110.
28. Wallach JD, Wang K, Zhang AD, et al. Updating insights into rosiglitazone and cardiovascular risk through shared data: individual patient and summary level. *BMJ.* 2020;368:17078.
29. Kirkham JJ, Dwan KM, Altman DG, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ.* 2010;340:c365.
30. Kirkham JJ, Riley RD, Williamson PR. A multivariate meta-analysis approach for reducing the impact of outcome reporting bias in systematic reviews. *Stat Med.* 2012;(31):2179–2195.
31. Zorzela L, Golder S, Liu Y, et al. Quality of reporting in systematic reviews of adverse events: systematic review. *BMJ.* 2014;348:f7668.
32. Saini P, Loke YK, Gamble C, et al. Selective reporting bias of harm outcomes within studies: findings from a cohort of systematic reviews. *BMJ.* 2014;349:g6501.
33. Golder S, Loke YK, Wright K, et al. Most systematic reviews of adverse effects did not include unpublished data. *J Clin Epidemiol.* 2016;77:125–133.
34. Gaissmaier W, Wegwarth O, Skopec D, et al. Numbers can be worth a thousand pictures: individual differences in understanding graphical and numerical representations of health-related information. *Health Psychol.* 2012;31(3):286–296.
35. Duke JD, Li X, Grannis SJ. Data visualization speeds review of potential adverse drug events in patients on multiple medications. *J Biomed Inform.* 2010;43(2):326–331.
36. Gotz D, Stavropoulos H. DecisionFlow: visual analytics for high-dimensional temporal event sequence data. *IEEE Trans Vis Comput Graph.* 2014;20(12):1783–1792.
37. Hakone A, Harrison L, Ottley A, et al. PROACT: iterative design of a patient-centered visualization for effective prostate cancer health risk communication. *IEEE Trans Vis Comput Graph.* 2017;23(1):601–610.
38. Galesic M, Garcia-Retamero R, Gigerenzer G. Using icon arrays to communicate medical risks : overcoming low numeracy. *Health Psychol.* 2009;28(2):210–216.
39. Cornelius V, Cro S, Phillips R. Advantages of visualisations to evaluate and communicate adverse event information in randomised controlled trials. *Trials.* 2020;21(1):1028–1012.
40. Phillips R, Hazell L, Sauzet O, et al. Analysis and reporting of adverse events in randomised controlled trials: a review. *BMJ.* 2019;9(2):e024537.
41. Gotz D, Wongsuphasawat K. Interactive intervention analysis. *AMIA Annu Symp Proc.* 2012;2012:274–280.
42. Perer A, Gotz D. Data-driven exploration of care plans for patients. In: *CHI '13 Extended Abstracts on Human Factors in Computing Systems.* New York, NY: Association for Computing Machinery; 2013;439–444. <https://doi.org/10.1145/2468356.2468434>.
43. Saket B, Endert A, Stasko J. Beyond usability and performance: a review of user experience-focused evaluations in visualization. In: *ACM International Conference Proceeding Series.* New York, NY: Association for Computing Machinery; 2016:133–142.
44. Wall E, Agnihotri M, Matzen L, et al. A heuristic approach to value-driven evaluation of visualizations. *IEEE Trans Vis Comput Graph.* 2019;25(1):491–500.
45. Stasko J. Value-driven evaluation of visualizations. In: *ACM International Conference Proceeding Series.* New York, NY: Association for Computing Machinery; 2014:46–53.
46. Chuang-Stein C, Xia HA. The practice of pre-marketing safety assessment in drug development. *J Biopharm Stat.* 2013;23(1):3–25.
47. Amit O, Heiberger R, Lane P. Graphical approaches to the analysis of safety data from clinical trials. *Pharm Stat.* 2008;7(1):20–35.
48. Zink RC, Wolfinger RD, Mann G. Summarizing the incidence of adverse events using volcano plots and time intervals. *Clin Trials.* 2013;10(3):398–406.
49. Zink RC, Marchenko O, Sanchez-Kam M, et al. Sources of safety data and statistical strategies for design and analysis: clinical trials. *Ther Innov Regul Sci.* 2018;52(2):141–158.
50. Karpefors M, Weatherall J. The tendril plot—a novel visual summary of the incidence, significance and temporal aspects of adverse events in clinical trials. *J Am Med Inform Assoc.* 2018;25(8):1069–1073.
51. 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 Med Res Methodol.* 2020;20(1):288–213.
52. Chazard E, Puech P, Gregoire M, et al. Using treemaps to represent medical data. *Stud Health Technol Inform.* 2006;124:522–527.
53. Izem R, Sanchez-Kam M, Ma H, et al. Sources of safety data and statistical strategies for design and analysis: postmarket surveillance. *Ther Innov Regul Sci.* 2018;52(2):159–169.
54. Mittelstädt S, Hao MC, Dayal U, et al. Advanced visual analytics interfaces for adverse drug event detection. In: *Proceedings of the Workshop on Advanced Visual Interfaces AVI.* New York, NY: Association for Computing Machinery; 2014:237–244.
55. Mayo-Wilson E, Golozar A, Cowley T, et al. Methods to identify and prioritize patient-centered outcomes for use in comparative effectiveness research. *Pilot Feasibility Stud.* 2018;4(1):95.
56. Vedula SS, Li T, Dickersin K. Differences in reporting of analyses in internal company documents versus published trial reports: comparisons in industry-sponsored trials in off-label uses of gabapentin. *PLoS Med.* 2013;10(1):e1001378.
57. Vedula SS, Bero L, Scherer RW, et al. Outcome reporting in industry-sponsored trials of gabapentin for off-label use. *N Engl J Med.* 2009;361(20):1963–1971.
58. Vedula SS, Goldman PS, Rona IJ, et al. Implementation of a publication strategy in the context of reporting biases. A case study based on new documents from Neurontin® litigation. *Trials.* 2012;13:1–13.
59. Mayo-Wilson E, Li T, Fusco N, et al. Practical guidance for using multiple data sources in systematic reviews and meta-analyses (with examples from the MUDS study). *Res Synth Methods.* 2017;9:2–12.
60. Mayo-Wilson E, Chen X, Qureshi R, et al. Restoring invisible and abandoned trials of gabapentin for neuropathic pain: a clinical and methodologic investigation (protocol). *BMJ.* 2021;11(6):e047785.
61. Karpefors M, Borini S, Edmondson-Jones M, Bijlsma H. Tendril - Compute and Display Tendril Plots. 2020. <https://cran.r-project.org/web/packages/Tendril/index.html>. Accessed September 1, 2022.

62. Wilkins D, Rudis B. Treemapify. 2.5.5. 2021. <https://wilko.org/treemapify/>. Accessed May 13, 2021.
63. Yeh S-T. Clinical adverse events data analysis and visualization [abstract PO10]. *Presented at The Pharmaceutical Industry SAS Users Group Conference, June 3–6, 2007, Denver, Colorado.*
64. Chuang-Stein C, Vu L, Chen W. Recent advancements in the analysis and presentation of safety data. *Ther Innov Regul Sci.* 2001;35(2):377–397.
65. 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.* 2020;10(6):e036875.
66. Siddiqui O. Statistical methods to analyze adverse events data of randomized clinical trials. *J Biopharm Stat.* 2009;19(5): 889–899.
67. Tsang R, Colley L, Lynd LD. Inadequate statistical power to detect clinically significant differences in adverse event rates in randomized controlled trials. *J Clin Epidemiol.* 2009;(62): 609–616.
68. Berry SM, Berry DA. Accounting for multiplicities in assessing drug safety: a three-level hierarchical mixture model. *Biometrics.* 2004;(60):418–426.
69. Chen C, Yu Y. Empirical studies of information visualization: a meta-analysis. *Int J Hum Comput.* 2000;53(5):851–866.
70. Plaisant C. The challenge of information visualization evaluation. *Proceedings of the Workshop on Advanced Visual Interfaces AVI.* New York, NY: Association for Computing Machinery; 2004:109–116. <https://doi.org/10.1145/989863.989880>.
71. Zhang Z, Wang B, Ahmed F, et al. The five Ws for information visualization with application to healthcare informatics. *IEEE Trans Vis Comput Graph.* 2013;19(11): 1895–1910.
72. Laurent J, Mcdevitt H. *Interactive Data Visualization of Adverse Events Clinical Trial Data with the D3.js Script Library.* Basel, Switzerland: Pharmaceutical Users Software Exchange; 2014. (Paper PP22). www.lexjansen.com/phuse/2014/pp/PP22.pdf. Accessed May 5, 2021.
73. Bailey R, Wildfire J, Bryant N, Rosanbalm S. *Monitoring Adverse Events Using an Interactive Web-Based Tool Immune Tolerance Network (ITN).* Durham, NC: Rho, Inc; 2015. https://rhoinc.github.io/publication-library/pubs/BioIT2015_AE_Explorer_Bailey.pdf. Accessed May 5, 2021.
74. Cornelius V, Phillips R. Improving the analysis of adverse event data in randomised controlled trials. *J Clin Epidemiol.* 2022;(144):185–192.