The Tendril Plot—a novel visual summary of the incidence, significance and temporal aspects of adverse events in clinical trials

Martin Karpefors[1] and James Weatherall[2]

[1]Advanced Analytics Centre, AstraZeneca R&D, Gothenburg, Sweden and [2]Advanced Analytics Centre, AstraZeneca R&D, Macclesfield, UK

Corresponding Author: Martin Karpefors, AstraZeneca R&D Gothenburg, Pepparedsleden 1, SE-431 83 Mölndal, Sweden. E-mail: martin.karpefors@astrazeneca.com

ABSTRACT

Background: In contrast to efficacy, safety hypotheses of clinical trials are not always pre-specified, and therefore, the safety interpretation work of a trial tends to be more exploratory, often reactive, and the analysis more statistically and graphically challenging.
Methods: We introduce a new means of visualizing the adverse event data across an entire clinical trial.
Results: The approach overcomes some of the current limitations of adverse event analysis and streamlines the way safety data can be explored, interpreted and analyzed. Using a phase II study, we describe and exemplify how the tendril plot effectively summarizes the time-resolved safety profile of two treatment arms in a single plot and how that can provide scientists with a trial safety overview that can support medical decision making.
Conclusion: To our knowledge, the tendril plot is the only way to graphically show important treatment differences with preserved temporal information, across an entire clinical trial, in a single view.

Key words: Patient safety, Visualization, Adverse event, Safety assessment, Clinical, Respiratory

INTRODUCTION

The understanding and communication of complex multidimensional data is dependent on good graphical representation. In fact, visualization is the key to effective, data-driven decision making in many areas, and therefore, it is important to keep inventing new visual designs, and seek to repurpose successful visualizations from other areas wherever possible. In this brief communication, we showcase how an artful visualization of Wikipedia discussions[1] can be transformed into a novel method of flagging potential safety signals in clinical trials.

It is important to capture and accurately record data on a wide range of adverse events (AEs), in order to understand the benefit-risk profile of a new medical intervention. In contrast to efficacy endpoints, safety findings are not as amenable to prespecification or prescription. While in practice, e.g., due to a known clinical interest of the disease or drug class, some safety endpoints are prespecified,

many AEs cannot be foreseen and may arise at any time during the clinical trial.[2] Therefore, it is of importance to find innovative ways to holistically illustrate safety, which show both the relative importance of risks and the temporal pattern of AEs in the study.

With regards to related prior work, we note that Amit et al.[3] presented a variety of useful ways of visualizing AEs in a trial, e.g., cumulative distribution of time-to-first AE and the complementary hazard function for a single AE, as well as a 2-panel combination showing paired incidence and relative risk for multiple AEs. No method is presented which shows the temporal structure of all trial AEs simultaneously. In the work of Zink et al.,[4] the volcano plot was introduced in a clinical safety context. This is a useful way of visualizing all the AEs in a clinical trial, such that they can be compared based on risk difference and *P*-value. However, of note is that this approach does not naturally lend itself to assessment of the tem-

poral profile of the AEs, other than by viewing a number of separate volcano plot snapshots, side-by-side.

In this paper, exemplified using data from a real clinical trial, we propose a completely new type of visualization that we call tendril plots. The tendril plot is a visualization that summarizes the overall AE profile in a study, clearly shows the AEs of major importance, and, in addition, illustrates the time course of reported AEs.

METHODS

Data

The data used to exemplify the method in this paper originates from a 52-week, randomized trial, Roflumilast Effect on Exacerbations in Patients on Dual [LABA/ICS] Therapy (RE^{2}SPOND, NCT01443845)[5] that evaluated the effect of roflumilast[6] in patients with chronic obstructive pulmonary disease. The population used in this paper consists of 2088 subjects, 1046 and 1042 subjects on the roflumilast and placebo arms, respectively.

AEs were coded according to the Medical Dictionary for Regulatory Activities (MedDRA) classification system. The preferred term (PT) was used as the key AE term for categorizing the AEs into branches in the tendril plot. All events, including recurrent events, were included in the analysis.

Tendril Algorithm

The tendril plot algorithm is inspired by a striking piece of work by Stefaner et al.,[1] in which their aim was to visualize the temporal flow of discussions about whether to keep or delete articles on Wikipedia.

We have taken a similar approach to summarize AEs in clinical trials. Figure 1 illustrates, on a single MedDRA PT (Back pain), how the tendril algorithm sequentially builds each branch, or tendril. The tendril is constructed from all back pain events and is grown upwards from the origin, which represents the start of the study. The distance in time from randomization to event runs along each branch. Changes to the direction of a tendril are dictated by which treatment arm the events occur in, which is highlighted with different colors in Figure 1. For each event on the active arm, the direction is tilted to the left, and for each event on the placebo arm, the direction is tilted to the right. The extent of tilting can be configured, for optimal visual exploration, and does not need to be the same for both arms. A larger value will make the tendrils separate more, which may be required to analyze small trials. In large trials with many events, the degree of tilting should be smaller. Furthermore, to prevent potential bias caused by unbalanced treatment allocation, the degree of tilting should preferentially reflect the proportion of subjects on the treatment arms. Note that all tendrils start their journey directly upwards, and, if AEs are balanced between arms, the tendril will always return to a vertical trajectory.

To build a tendril, the algorithm steps for each AE are the following:

- Sort the events according to time since randomization.
- Calculate the magnitudes of the vectors as the time between subsequent events. For an event occurring at the same time as the previous event, the magnitude will be zero.
- Calculate the angle of the vectors. For each vector the angle is the cumulative sum of all angles up to that event. The angle is negative (clockwise rotation) for events on the placebo arm and positive (counter-clockwise rotation) for events on the active arm. Zero-magnitude vectors will still contribute to angular changes. Thus, if 3 events, 2 on placebo and one on active, occur at the same time, the net effect is a 1 unit clockwise rotation.
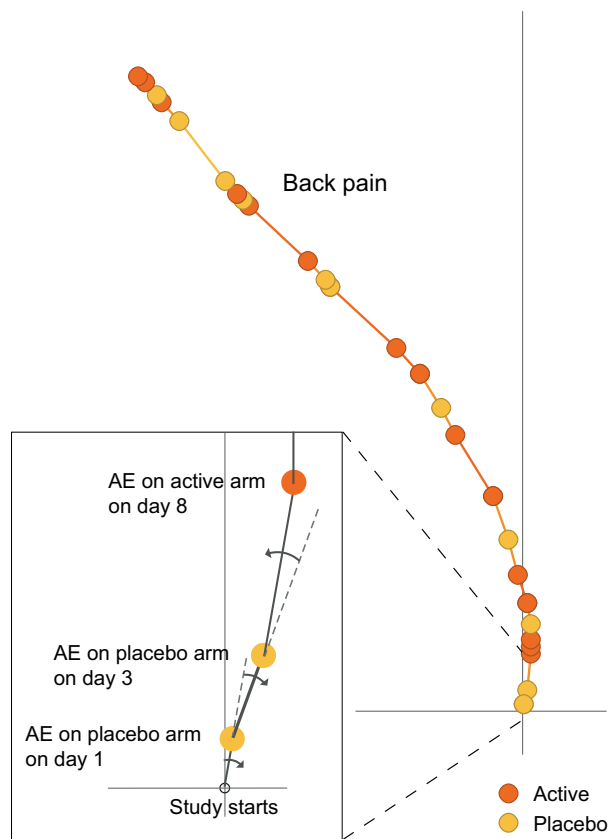

Figure 1. The Tendril plot algorithm concept illustrated on a portion of a single preferred term: Back pain. The inset zooms in on the first 3 events to demonstrate how events on placebo tilt the tendril to the right and events on active tilt the tendril to the left. To emphasize this step in the algorithm, in this plot, the events are colored by treatment arm; yellow and brown circles for placebo and active arm, respectively. The distance between points are proportional to time between events. Point size is not linked to any information.

- Add the vectors together cumulatively, i.e., the next vector in time starts at the end of the previous vector in time. The resulting sequence of vectors constitutes the tendril for that AE.

The tendril plot algorithm was implemented as an R package called Tendril (developed in R, Version 3.2.4)[7] which is available at Github.[8]

The tendril plot for the RE^{2}SPOND trial is shown in Figure 2. Each tendril in the plot is represented by a MedDRA PT, having at least 10 incidences in at least one of the two treatment arms. From trial-and-error testing and the 1:1 treatment allocation, the tilting angle per event was chosen to be ±4° for events on the placebo and roflumilast treatment arm, respectively. Event coloring in Figure 2 is based on false discovery rate (FDR)-adjusted *P*-values (see statistical considerations section below) and point sizes are proportional to the total number of PT events per tendril, including recurrent events.

Interactivity

To exploit the tendril plot to its fullest, an interactive environment is required. In this work we have used Tibco Spotfire (Version 7.0.1). The dataset generated by the R algorithm was imported into Spotfire, where a scatterplot, showing the tendril coordinates connected by AE, constitutes the basic view. At this point it is possible to pursue interactive exploration of the data through filtering, e.g.,
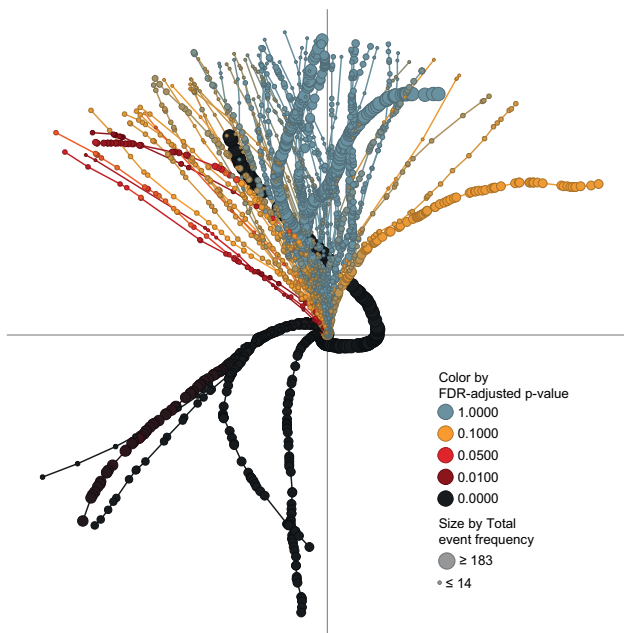
Figure 2. A tendril plot of all AEs in the RE^{2}SPOND trial. Each MedDRA PT is represented by a line (tendril) and each point is an event. Since time runs along each tendril, it is the shape that carries the important information, rather than the x and y coordinates. An event on the roflumilast treatment arm will tilt tendril direction to the left, and an event on the placebo arm will tilt tendril direction to the right. Point size is indicative of the total number of events for the type of AE in both treatment arms in the trial. The FDR adjusted Pearson's chi-squared P-value in each point is mapped onto a continuous color gradient.

focusing in on AEs of particular interest; hovering, getting contextual information via popups; remapping, changing colors, and other plot characteristics to understand different aspects of the data; and linking, that is, creating customized detailed visualizations. For example, the plot characteristics in Figures 1 and 2 are configured to highlight different aspects of the visualization.

Statistical Considerations

Since the range of possible AE outcomes is large and unpredictable, statistical testing cannot be prespecified in the same way as for efficacy outcomes. For large clinical trials there can be thousands of events. A classical Bonferroni correction for multiple testing would result in a very small P-value threshold, which, for the purpose of flagging AEs requiring further investigation, would be too conservative. Instead, we have made use of the more relaxed FDR as proposed by Benjamini and Hochberg.[9] FDR-adjusted P-values give a better balance between type I error and power, and decrease the amount of false negatives as compared to the stricter Bonferroni scheme. In the algorithm, for each tendril and cumulatively for every time point of events, we calculate the Pearson's chi-squared P-value for the hypothesis that the treatment arms have the same proportions of events up to that event. The P-values are then FDR adjusted and mapped onto a color gradient as shown in Figure 2.

RESULTS

In the RE^{2}SPOND study, 3085 AEs were reported between randomization and the end of the study, for AEs having at least 10 incidences in at least 1 of the 2 treatment arms. The AEs are distributed on

16 System Organ Classes (SOCs), 34 High Level Group Terms, and 69 PTs according to the MedDRA hierarchy.

The tendril plot summarizing the RE^{2}SPOND trial is shown in Figure 2. It has 69 branches or tendrils, 1 for each PT. There are 6 tendrils (namely diarrhea, nausea, weight decreased, insomnia, headache, and decreased appetite) that noticeably differ from the others. In those 6 AEs, there is an early leftward bend in the propagation, indicative of an unbalanced proportion of events in the treatment arm early on in the trial. The early and sustained imbalance is also highlighted by the black coloring throughout. For highly unbalanced events like diarrhea, the net difference between numbers of events is >90, thus exceeding a full 360° rotation (at a 4° rotation per event).

Nausea, decreased weight, insomnia, and headache are only transiently unbalanced, evolving into a straight balanced line shortly after start of the trial. Using the interactive features of Spotfire, we can explore details of the tendril plot. Hovering over the plot, as shown in Figure 3A, provides information including PT name, incidence of events on the different treatment arms, and time since randomization of events. Selection of a single PT in the filter menu, e.g., nausea (see Figure 3B), will momentarily remove all other AEs. Zooming in on the early part of the tendril and selecting the events occurring during the first 7 weeks will, in a linked bar chart visualization (see inset in Figure 3D), reveal that there are 48 and 8 cases of nausea in the roflumilast and placebo arms, respectively. However, in the remaining 45 weeks of the trial, the picture is more balanced: 26 occurrences of nausea on roflumilast as compared to 24 on placebo. Diarrhea, the most common AE and by far the most extreme tendril, tends to balance out in the later part of the study.

There are many ways of filtering the data. In Figure 3C a range filter was used to interactively limit the view to only those tendrils having 100 events or more in total. The restriction resulted in a tendril plot containing the 8 most common AE branches. In Figure 3E the tendrils are remapped into panels representing the SOCs to reveal any potential imbalances on the body organ level.

In contrast to the outstanding and unbalanced AEs discussed so far, the majority of the reported AEs had a similar distribution of events on both of the treatment arms, and therefore will follow a near vertical trajectory. To simplify the interpretation, especially in the borderline regions, colors can be utilized to flag observations of interest.

DISCUSSION

We have described and exemplified a new approach to the interactive visual interpretation of AEs across an entire clinical trial. The approach, the tendril plot, includes all the features of previous similar methods, and also adds a level of interactivity and temporality absent from those prior methods. When trials have a large volume or variety of AE data, this approach may be particularly helpful, in order to holistically assess the risk profile of a new therapy.

Using the data from a real clinical trial to exemplify the approach, we demonstrated that AEs which are indicated as common in the drug label, were also highlighted in the tendril plot. We also demonstrated the versatility of the tendril plot, via informative snapshots of the interactive tool, indicating the dynamism and flexibility that an analyst can expect.

There are some limitations of the tendril plot, as with all methods. Firstly, it is not easy to move beyond pairwise treatment comparisons, without creating an entirely new plot. Secondly, if a patient is taken off treatment following a particular AE, that AE can appear to be transient when it is not. A useful addition to add
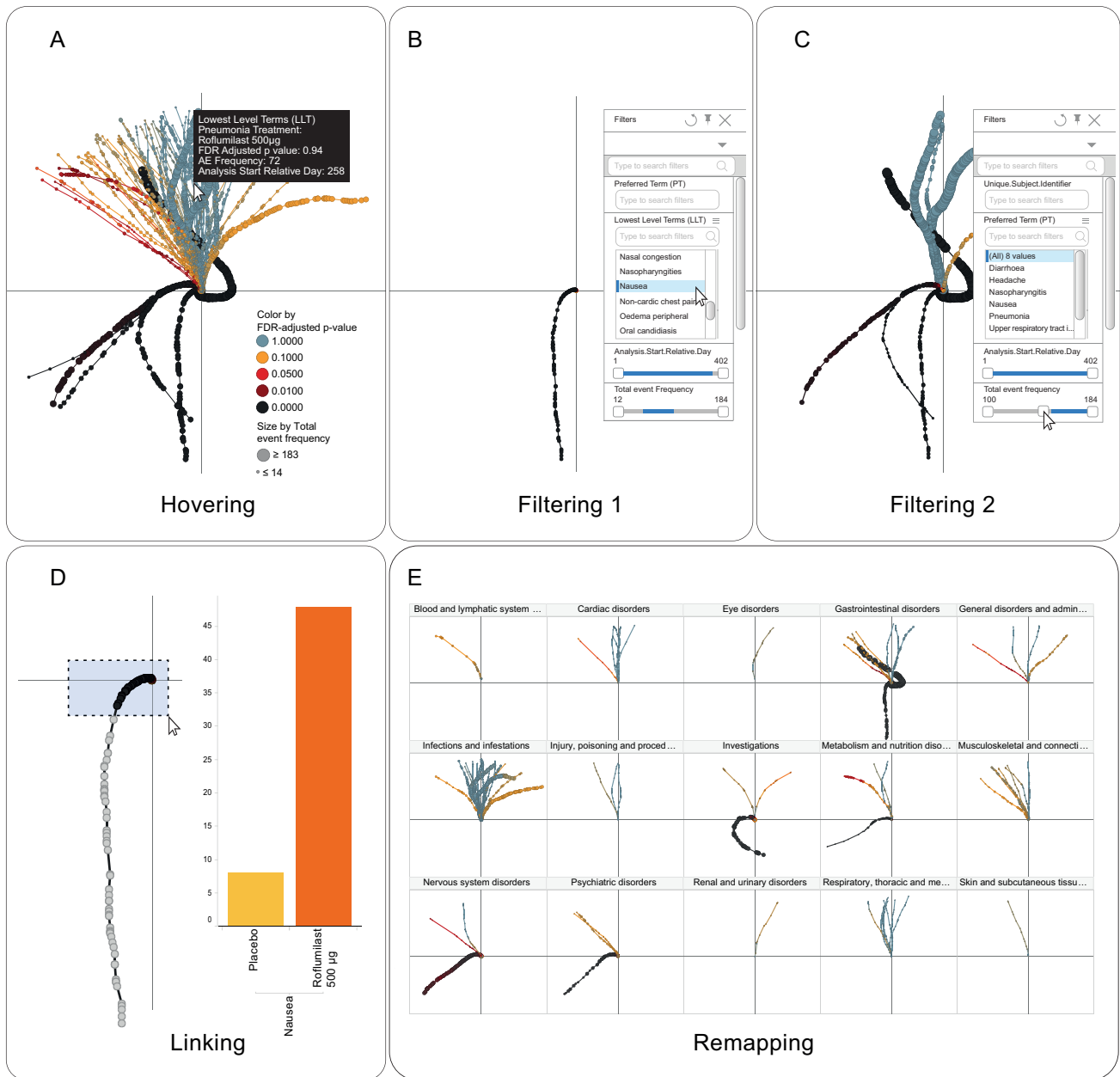
Figure 3. Panel showing examples of interactivity. For example (**A**) hovering over an event gives detailed information about type of AE, AE frequency, etc.; selecting a specific PT in (**B**); filtering to a subset of PTs with a total event frequency ≥100 in (**C**); linking selected events to additional information in (**D**); and recasting into a trellis configuration, for assessment by SOC, in (**E**).

further robustness to the interpretation of the tendril plot would be to make allowance for both baseline and ongoing imbalances between trial arms. For instance, this could be done by calibrating the bend severity in proportion with patients remaining at risk in each arm.

It should be noted that there is an inherent trade-off in the tendril plot, between aesthetic appeal and mathematical convention, both being important in our view. For instance, in the current implementation of the plot, there is no strict interpretation of what the "x" and "y" axes denote. Alternative implementations could be explored in future work, e.g., where study days are measured along x and AE frequency along y.

We would point out that, while tendril plots are an accessible way to assess AEs across an entire clinical trial, when single AEs are

identified for further investigation, more formal statistical testing should still be pursued.

Compared with previously published methods, the tendril plot provides a more compact and temporal picture of AEs across the whole trial. Volcano plots provide a nice holistic summary of trial AEs, but require multiple such plots for a genuine representation of time course. For both volcano and tendril plots, moving beyond pairwise treatment comparisons is not straightforward. In the case of the tendril plot, we would propose to make use of interactive software, allowing the analyst to toggle between different pairs of treatments, in order to assess each in turn.

Furthermore, an interactive environment allows informative filtering and recasting of data, for instance into a multipanel view of

plots, representing the MedDRA SOCs. Managing the visualization of large clinical trial data sets can be challenging, but when provided, it gives a powerful means for medical hypothesis generation and decision-making.

Finally, this novel concept of utilizing the branch *direction* for carrying important information could be applied in other areas. In observational data sets, such as electronic health records, there may be multiple applications: tracking prescribing patterns between pairs of drugs in the same class; profiling symptoms and putative side effects between drug pairs; examining imbalances of events between paired cohorts, such as young vs old, men vs women, or those with a particular genetic mutation vs no mutation.

FUNDING

COMPETING INTERESTS

The authors are AstraZeneca employees with no competing interests to declare.

CONTRIBUTORS

MK and JW both contributed to the intellectual content of conception and design. MK coded the algorithm and conducted the study. MK and JW jointly wrote the manuscript.

REFERENCES

1. Stefaner M, Taraborelli D, Ciampaglia GL. *Notabilia - Visualizing Deletion Discussions on Wikipedia*. 2010. http://www.notabilia.net. Accessed September 12, 2017.
2. Crowe BJ, Xia HA, Berlin JA, *et al*. Recommendations for safety planning, data collection, evaluation and reporting during drug, biologic and vaccine development: a report of the safety planning, evaluation, and reporting team. *Clin Trials* 2009; 6: 430–440.
3. Amit O, Heiberger RM, Lane PW. Graphical approaches to the analysis of safety data from clinical trials. *Pharm Stat* 2008; 7: 20–35.
4. Zink RC, Wolfinger RD, Mann G. Summarizing the incidence of adverse events using volcano plots and time intervals. *Clin Trials* 2013; 10: 398–406.
5. Martinez FJ, Rabe KF, Sethi S, *et al*. Effect of roflumilast and inhaled corticosteroid/long-acting b2-agonist on chronic obstructive pulmonary disease exacerbations (RE^{2}SPOND) - a randomized clinical trial. *Am J Respir Crit Care Med* 2016; 194: 559–567.
6. Calverley PM, Rabe KF, Goehring UM, *et al*. Roflumilast in symptomatic chronic obstructive pulmonary disease: two randomised clinical trials. *Lancet* 2009; 374: 685–694.
7. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria; 2015. https://www.R-project.org. Accessed November 13, 2017.
8. Karpefors M. Tendril R package repository. https://github.com/Karpefors/Tendril. Accessed November 13, 2017.
9. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B* 1995; 57: 289–300.