

Medical Data Review and Exploratory Data Analysis using Data Visualization

VINOD KERAI, ROCHE, WELWYN, UKINTRODUCTION

Drug Development has drastically changed in the last few decades. There are huge costs involved in bringing a new drug to market and Pharma companies are under pressure to deliver innovative products quicker, with reduced budgets whilst maintaining regulatory compliance.

This is coupled with an entirely new problem; in recent years data generation has grown exponentially. Driven by the increased use of electronic medical records, technological advances in genomic sequencing and monitoring devices, the challenge of handling “Big Data” is becoming more apparent in the Pharma industry. Not only is the sheer volume of data growing, but the variance in complexity and the depth means that there is now an opportunity, and in some ways an obligation, for clinical scientists to mine this data to gain insights and understandings that have never been possible previously. Using this data to understanding trends and identify potential issues could mean the difference between a successful and sustainable launch and a drug being late to market. So making sense of all this ‘Big Data’ is an important and significant challenge.

This vast amount of data needs to be translated into information, and this information must be meaningful and impactful. It needs to be delivered in an easily-digestible format that allows the key stakeholders to make informed decisions swiftly and efficiently.

DATA VISUALIZATION

Data Visualization is the visual representation of data in order to enhance the analysis of large quantities of data and translating it into meaningful information. It represents the data in a way to engage the user as well as helping to communicate complex ideas quicker in a way to enable the end user (e.g. scientists) to discover patterns that might otherwise be hard to see in simple tables or listings.

PhUSE 2012

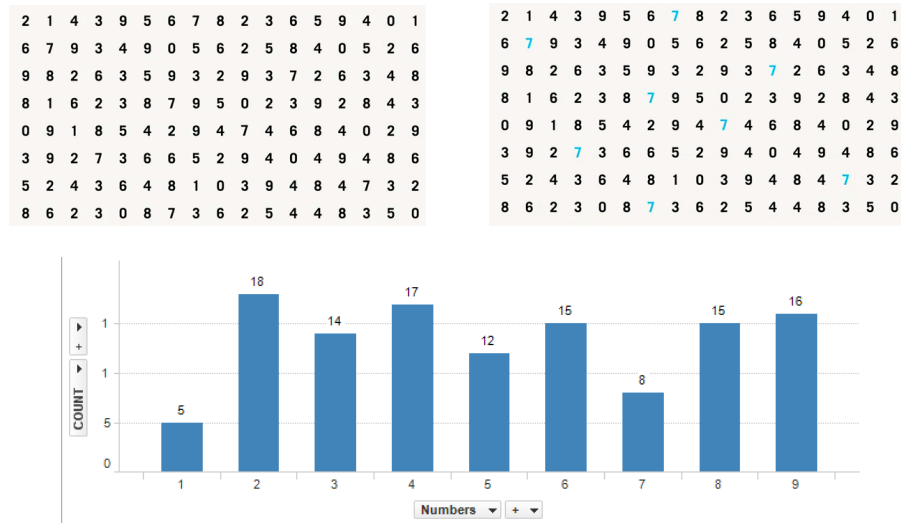


Fig 1. A simple piece of data visualization, count the number of 7s in both listings above, the simple addition of color highlights the numbers of interest, or a simple bar chart provides much more information visually. Data visualizations offer ways to find trends and correlations that can lead to important discoveries. Visualizations allow to understand and process enormous amount of information quickly because it is all represented in a single image or animation. There is rarely a single visualization that answers all questions. Instead, the ability to generate appropriate visualizations quickly is the key.

DATA VISUALIZATION IN ASSISTING MEDICAL DATA REVIEW

Clinical trials are now producing vast amounts of data and of increased complexity and variety. A new method of comprehending data is clearly needed in order to turn this data into useful information.

Roche is encouraging project teams to increase their use of data visualization via various tools, including Tibco Spotfire, to improve both efficiency and quality of review/analysis activities.

In order to drive science by implementing smart solutions for data delivery and data visualizations for different study types, a new dedicated DPS (Data Provision Specialist) role has been created which will allow science to focus on the review of safety data while making best use of the available resources and tools. The DPS is responsible for providing dedicated, desk-side and real-time technical support to colleagues in clinical science as they review clinical data for safety purposes. A key skill that the DPS has is the ability to understand both the science behind the data as well as the technical aspects of the data structures and the logical connections between the different data types and locations. The scientific understanding enables the DPS to translate science questions into evaluation steps and guide scientists through the question and answer cycle of exploratory data analysis. Having this technically minded dedicated resource allows the scientists to focus on the trends and the 'information' provided rather than having to mine through the vast amounts of 'big data'. The primary objective of this partnership is to effectively and quickly understand the safety medical data review needs of science and to translate these into meaningful and useful data visualizations. Scientists have quick access to specifically created data displays, allowing them to make more accurate decisions. The visualizations can be quickly tailored to address specific questions and to identify of trends and signals.

The scientists will usually have a specific section of the data they will want to look at and often have theories that they would like to confirm or monitor for the safety of subjects in clinical trials. For in a large oncology trial there typically a large number of adverse events and usually of a more complex type

then in studies in other of other therapeutic areas. Often an adverse event can last the duration of the study (which can be several months in some cases) and is monitored throughout. Viewing this large amount of information is becoming more and more difficult using the conventional method of tables and listings. And when the scientist wants to correlate the adverse event (AE) with other information from the trial, this makes an already difficult task even harder!

In a recent oncology trial at Roche, such a problem started to arise: large amounts of complex data with an increased variety were being generated. As can be seen in Figure 2 listings were created and even by focusing on one particular adverse event, 37 rows for AE intensity changes, 2,000+ lab results, and 8 AE treatments were created. Overall there are 2100 different AEs across 155 subjects in the study right now. This was causing the science team great challenge about viewing the data simultaneously and was cumbersome to navigate from AE to AE and subject to subject.

An interactive dashboard was created in Spotfire with a close collaboration between the DPS and the clinical scientists to ensure all their information needs were being met. A key focus was to enable the data to be easily navigable by the scientists while the DPS ensured that all the technical aspects were correctly aligned. A total of 14 visualizations in one single dashboard were created initially to show all the different aspects of the AE data. Figure 2 shows one such output. It monitors any particular lab value to see if it increases with the increase in intensity of the AE. This is a fully interactive graph that aligns all the different data with trends highlighted by colours and lines. Programmed functions ensure easily navigation via drop down lists. The scientist used the drop down lists to select parameters such as subject number, lab parameter, adverse event of interest. Again, the key focus is to ensure that the scientist are able to work with the data and the information rather than working on the technical aspects of the tool. This overcomes one of the main stumbling blocks in the uptake of these new ways of working.

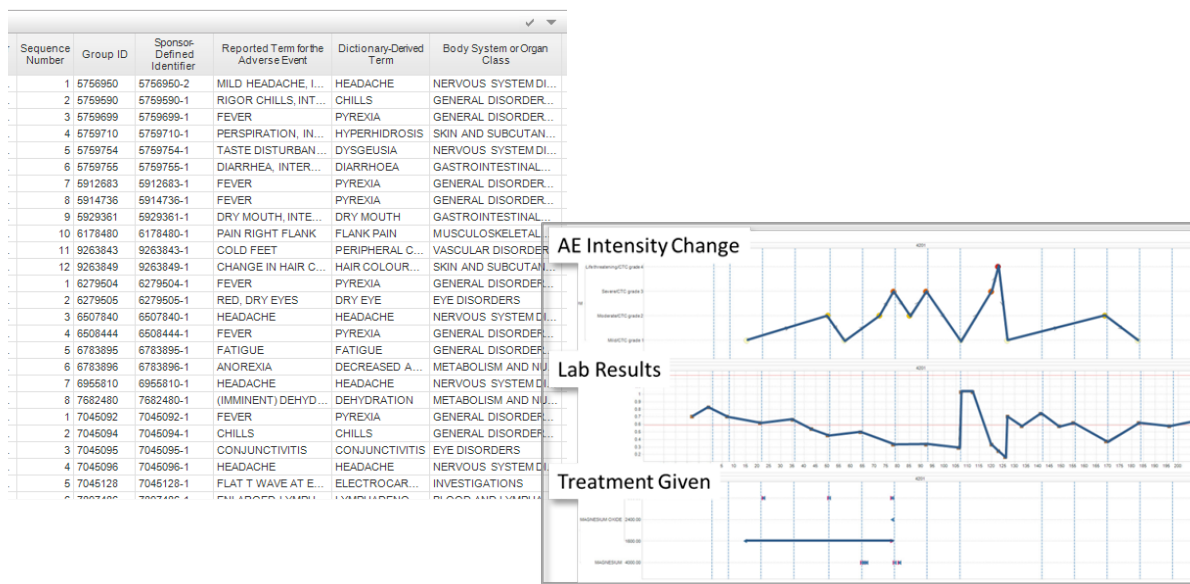


Fig 2. A simple piece of data visualization for Medical Data Review, (Left) A listings view (Right) An interactive graph within a dashboard of 14 different visualization outputs.

The monitoring of RECIST tumor data is another example of very successful use of data visualization. RECIST data is collected on the eCRF in a very different way to how it is reported in the SDTM format. It

PhUSE 2012

is collected in 5 different eforms on the eCRF while the data is then split into 3 different datasets in SDTM. Once again the problem of big data exists even when only looking at one type of data. The technical aspects of this data coupled with the need to calculate a large number of derived values (change from Nadir/ change from baseline) make using raw tumor data extremely difficult.

Previously this data was manually transcribed from the eCRF into Microsoft Excel and calculations done manually by the scientists. This was proving to be a time consuming process and often meant reprocessing and rechecking the data as it was constantly changing due to data management processes cleaning and correcting the data.

Once again a close collaboration between the DPS and the science teams lead to the creation of a bespoke dashboard to collate all this data into one interactive output. Links were created to the SDTM domains which would be updated daily, based on the live eCRF data, when coupled with the data visualization capabilities and calculation functions within Spotfire and the knowledge of the DPS, meant the clinical scientists were able to see the live data with no overhead on recalculations and rechecking of the data. Figure 3. shows one such output that visualizes an individual subject's tumor size changes both from baseline and Nadir and highlights the trend in drug response on the overall response of the tumor sizes.

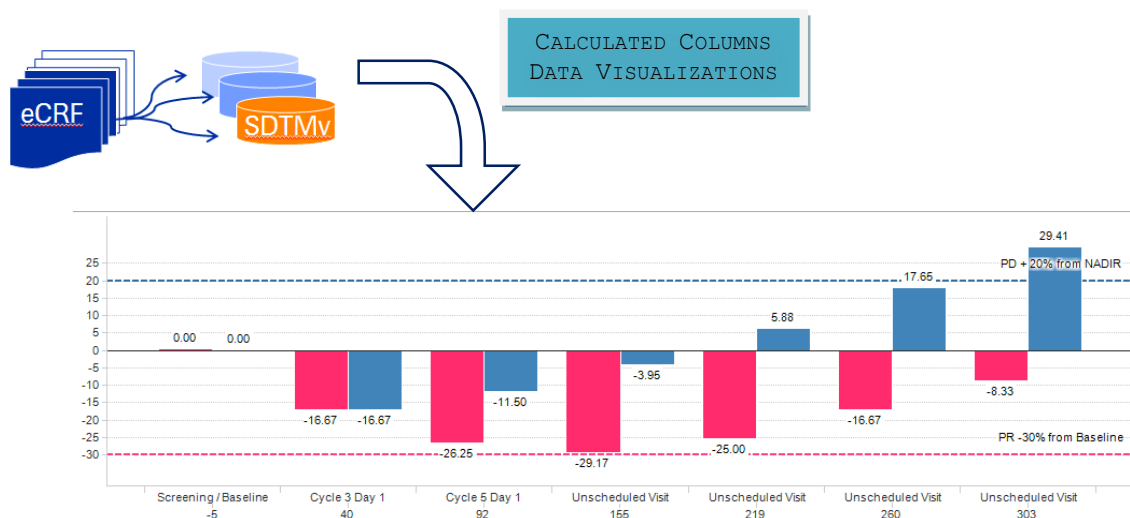


Fig 3. RECIST Responses for a subject throughout the study. A large number of calculations need to be done in order to calculate % change from baseline and % change from NADIR, previously manually done in spreadsheets and tables, now instantly calculated.

Another example of data visualization having a positive influence is the way that correlations and efficacy signals can be investigated on a real time basis. This is a pioneering new way to carry out exploratory data analysis. Medical data review is a planned and structured approach to looking at the data in a clinical trial, going through question and answer cycles while interrogating the data. Exploratory data analysis is an ad-hoc real time approach to looking at data to investigate potential signals, correlations or trends that may potentially exist in the study. This has not been possible before due to a combination of reasons. With multiple hand overs of the dataflow due to legacy processes and organizational structures, it would have taken weeks/months for programmers to generate specific outputs based on inflexible and restrictive standard reporting outputs.

Due to the fact that a DPS partnership with scientists is established throughout the entire lifecycle of a study, these handovers are now almost nonexistent and data is available for exploration in a much shorter timeframe.

The fully interactive capabilities of the software, coupled with DPS knowledge of the study, therapeutic area and technical knowledge in the data structures, it is now possible for a scientist's hypotheses of trends and signals to be confirmed in a matter of hours rather than days or weeks.

One such case was seen recently in a study and is shown in Figure 4. The investigator and clinical scientist were suspecting that large drops in lymphocyte counts were causing AEs with higher intensities. After discussing with the DPS and providing a simple sketch (left) of what was needed to be shown, the output (right) was quickly created. Once created it can be monitored constantly throughout the study lifecycle and data is now updated automatically with very little rework.

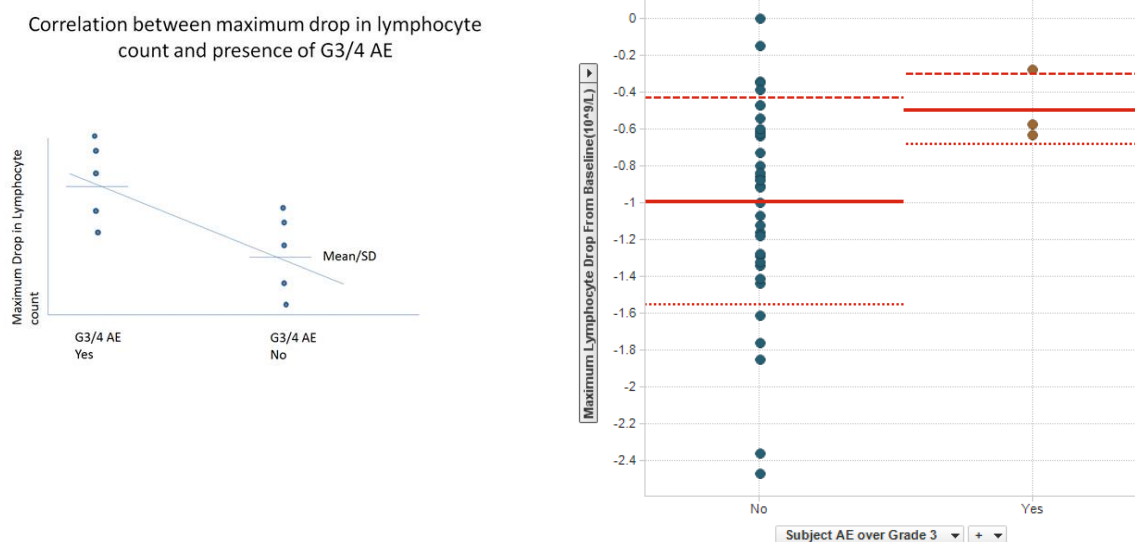


Fig 4. A piece of data visualization for identifying correlations and safety signals, maximum drop from baseline in lymphocyte counts against AEs over CTC Grade 3.

PhUSE 2012

CONCLUSION

This new way of working, and the collaborations being formed between the DPS and the Science groups, is now flourishing at Roche with almost 25% of studies using these new tools and methods. There has been great feedback from both sides and the growth in demand shows how this is proving to be a better and more efficient way of looking at the data in a clinical study.

It enables the right people to concentrate on their specialties. Science can concentrate on looking at their data and applying their clinical judgment. The DPS can focus on the technical aspects to bring the data and the information closer to the science teams enabling them to make their decisions more accurately and on a timelier manner.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Vinod Kerai
Roche Products Ltd
6 Falcon Way,
Shire Park,
Welwyn Garden City,
AL7 1TW
UK

Tel. +44 (0)1707 366563
Fax +44 (0)1707 384118
vinod.kerai@roche.com

Brand and product names are trademarks of their respective companies.