

Translating eClinical Data – The Art and Science of Data

Nikki Dowlman, PhD
Perceptive Informatics, Nottingham, UK
nikki.dowlman@perceptive.com



ABSTRACT

With the increasing complexity of studies in their design, technologies employed and sheer size, it is more critical than ever to quickly assimilate large amounts of data to facilitate the conduct of a study, identify action areas, and share best practices. As a consequence, many sponsors are seeking to amalgamate data in a warehouse to minimise the hunt for data in disparate places. This paper will explore best practice methodologies that can be employed to ensure intuitive and informative display of data, as well as processes for ensuring useful metrics are identified. We will review the Gestalt Principles of visual perception and the importance of User Interface (UI) design for data display. Case studies of both good and bad display of information from other industries will be used to provide illustrative discussion points. These principles will then be applied to the data in the clinical trial industry.

INTRODUCTION

Business Intelligence (BI) has been around for a long time. In a 1958 article, Hans Peter Luhn cited a Webster's dictionary definition of BI: "the ability to apprehend the interrelationships of presented facts in such a way as to guide action towards a desired goal." But it was not until the late 1990's that it became the 'next big thing' for companies worldwide. Historically BI belonged in the environs of the boardroom, providing information on company performance for executive decision making. Recent years have seen its importance expand into the operational realm to help manage a company's Key Performance Indicators (KPIs).

As clinical studies increase in complexity in a myriad of ways, a key question often asked is, *"Is our ability to create complexity increasing faster than our ability to understand complexity?"*

This is an exciting time to be involved in the reporting of data and metrics on the performance of a clinical study. However, it is important to understand the principles of visual presentation of data to ensure that the information is accessible, actionable, not misleading and ultimately valuable for the end consumer of information.

WASHED AWAY IN A SEA OF DATA

Paul Allen (co-founder of Microsoft) was quoted as saying, "For me, goals and daily metrics are the key to keeping me focused. If I don't have access to the right stats every day, it is so easy for me to move on mentally to the next thing. But if I have quick access to key metrics every day, my creativity stays within certain bounds—my ideas all center on how to achieve our goals." Having the right information readily accessible especially for data-intensive clinical trials at the right time and in a format that is easy to understand helps focus on what is important. Without focus it is easy to become distracted and dwell on issues that do not have a positive impact on objectives. Keeping these basic principles in mind when defining metrics and designing how to display them is essential in delivering metrics that are valued by the user.

WHAT IS A GOOD METRIC?

A common approach to evaluating the effectiveness of metrics is that they should fulfill 4 criteria. Metrics should be:

1. Simple
2. Small
3. Actionable
4. Used

Metrics should be easy for the user to interpret. They should not require effort on the part of the consumer to understand what they are trying to say. A metric should not try to measure more than one thing. The consumer should be able to action the

PhUSE 2011

metric, meaning if an action is required to effect whatever the metric is measuring, the consumer should be empowered to do so. After all, if the metric shows things are going wrong but the consumer is powerless to do anything, there is little point in informing them. This means the consumer must use metrics. Simply because something is measured and could be actioned, does not mean it always is.

Identifying and understanding the consumer is central to identifying appropriate metrics. It is easy to create metrics that have been driven by what information is available in isolation from the consumer, thereby lacking focus and ultimately delivering inappropriate tools for the job. Instead of this approach, it is better to start with a different question. For example, *'what are you responsible for making happen?'* In this way you can ensure that every user gets the right information for the right task at hand.

A multitude of roles are involved in a clinical trial such as Clinical Study Leader, Clinical Supply Managers, Statisticians, Country Managers, CRAs/Monitors, and Site Personnel. In product management terms, these roles are considered User Personas. A clear understanding of these roles and the activities they are required to perform by asking the question *'what problems are you trying to manage?'* provides the framework for identifying the information they need to perform their jobs.

Metrics should be about the past, the present and the future. For example, *'what is the current state?'*, *'how did I get here?'* and *'where will I be in the future?'*. Table 1 illustrates some of the thinking that goes into creating metrics around site activation that are likely to be of interest to a Study or Country Manager.

Table 1

Question I need to answer	Answers that can be displayed as a metric
How many sites do I have active?	Total number of sites that are active.
How many should I have active?	Total number of sites that are NOT active but should be based on planned activation date.
How big is the discrepancy?	• What is my current activation rate? • Is that greater or less than my planned activation rate? • Is the discrepancy getting bigger or smaller compared to historic information?
How do I get back on target?	• Provide 'model' activation rate to get back on target. • Is this target rate more or less than it was in the month previous?
If I don't get back on target, when will I finish?	• Based on current rate of activation, the predicted completion date for the activation process. • Is this predicted date getting closer or further away?
Where is it going right/wrong?	There are a number of trends that could be analyzed. Two examples include: • Display activation through time such as a line chart to determine if there are any trends that can be observed (e.g., does a national vacation slow activation more than expected). • Look at activation rates (e.g., by Country Manager or Monitor). Is one performing better than another? Is there best practice that can be shared?

Also important is understanding the lifecycle of a clinical study which, at its broadest, can be considered inception, design, start-up, subject recruitment, study conduct, close out, and submission. These lifecycle steps should be used to influence when information is presented and how its prominence may change over time, depending on the study status. Delivering information that is relevant to the user in alignment with business processes will help ensure that metrics are actionable and used.

PhUSE 2011

There are generally 4 types of metrics that one could measure in the clinical trial industry:

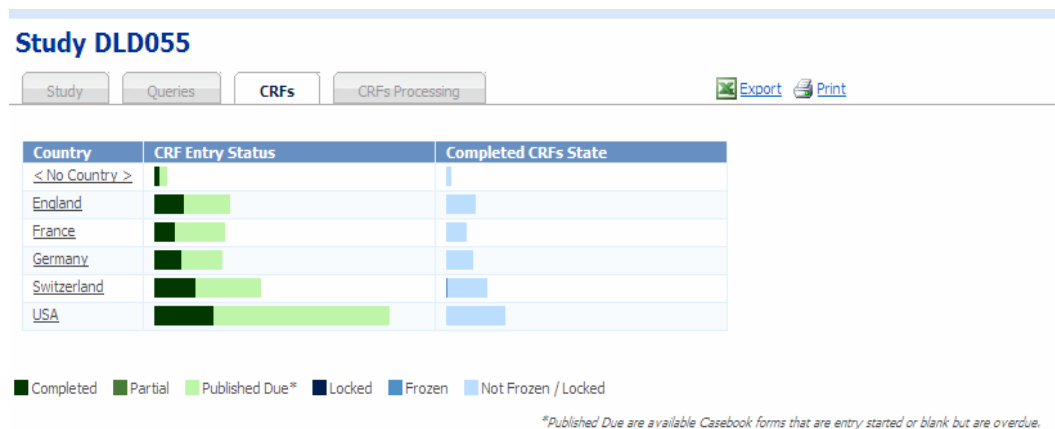
1. Cycle time – how long an activity takes
2. Timeliness – whether an activity was completed on time
3. Quality – a measure of the amount of errors or rework required
4. Cost – a measure of incurred resources

Anecdotally, it would appear that while there is much focus on metrics addressing both cycle time and timeliness, there are fewer that monitor quality and cost within the clinical trial industry. These four elements need to be in balance to ensure that doing something on time does not result in a lot of rework along the way. One of the metrics that is often discussed in Electronic Data Capture (EDC) solutions is the close out rate of queries (i.e., the time taken from when a query is raised until it is resolved). This is a cycle time metric. Although important in a trial, this is measuring rework (i.e., a lack of quality). This is an example of an ineffective resource investment as it does not materially contribute to the outcome of the trial. Conversely, focusing effort on reducing the number of queries raised has twice the impact as it improves quality AND reduces the amount of resources required to do the rework. Focusing in this area is an effective use of resource as it materially contributes to the outcome of the study.

CONSOLIDATING METRICS

One of the challenges in the clinical trial industry is the sheer number of different technologies each supporting a different part of the clinical trial process, such as Randomization and Trial Supply Management tools (RTSM using Interactive Voice or Web Response), Electronic Patient-Reported Outcomes (ePRO), EDC, and Clinical Trials Management Systems (CTMS), to name a few. Consequently, there is much discussion around data warehouses, integrations, enterprise service buses, middleware and ETL (Extract, Transform, Load) processes. It is safe to assume that the consumer of the information does not care about the technology solution that sits underneath the information. The value is in the synergies created by consolidating the data and identifying new metrics. For example, the knowledge that a Case Report Form is expected in the EDC system because a visit has occurred in the RTSM solution, helps manage activities and processes by filling information gaps (Figure 1).

Figure 1



VISUAL PRESENTATION OF METRICS

Having defined the content of your metrics and data reports, a key step in the process is to appropriately display the information. Our human eyes have one of the highest bandwidths to provide information to our brain. Edward Tufte is considered one of the fathers of information design starting his work in the 1970's. However, it has provenance going back to the 1800's with John Snow's spot maps of a cholera outbreak in London in the 1850's, leading to the discovery that the disease is spread by contaminated water, as well as Charles Joseph Minard's 1869 flow chart on Napoleon's Russian campaign of 1812. Although the concept of metrics has been around for a long time, it has not been applied to the clinical trial industry with significant rigour. In fact, the pharmaceutical industry is behind other industries where there are many examples of metrics use.

PhUSE 2011

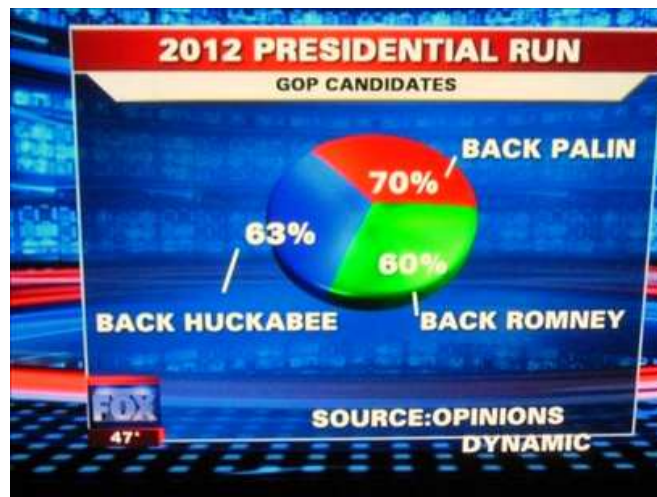
The industry is in the process of bringing more robust metrics, enabled by technology, to bear to make trials more efficient and effective – and importantly, to make better decisions faster about development. There are many best practice methodologies that can be employed to ensure intuitive and informative display of data via eClinical solutions.

The Gestalt Principles of visual perception as well as the importance of User Interface (UI) design for data display are instrumental in helping users absorb as much information as possible as quickly as possible. It is also important to select the correct visual mechanism to display the information. For example, which of the bar chart, line graph or table options is the most appropriate way to display a particular piece of information?

Seminal papers published as far back as 1956 illustrate that we tend to remember 7 +/- 2 bits of information at any one time (George A. Miller 1956 “The Magical Number Seven, Plus or Minus Two”, Psychological Review). This remains one of the key data presentation principles for delivering meaningful data to users. There are limits to the number of ‘chunks’ of information that anyone can remember readily and these differ for words, numbers and pictures. You may remember the principle from a childhood party game where a number of objects were placed on a tray before being covered. When one item was removed and the tray uncovered, the winner was the one who correctly identified the missing object. The more objects, the harder it became to correctly identify the removed item.

Visual presentation of data is a way of encoding large amounts of information into a ‘chunk’ that the consumer can easily interpret. This is because the brain is very adept at spotting patterns in what it sees. However, care is needed as this can mislead unintentionally or even, on occasion, deliberately. Figure 2 shows an example of a chart entitled, ‘2012 Presidential Run’. This is an example of a poorly designed and potentially misleading chart. Firstly, a pie chart indicates a part to whole relationship (i.e., one or more objects is part of another object). This means the segments should sum to 100% but, the segments in the example sum to 193%. Secondly, the values are required on the chart in order for the consumer to actually be informed as to which segment is bigger. This is because the eye is not optimal at comparing either a 2D area or the angles formed by the slices. Therefore, this particular chart adds no value. Finally, reviewing the question asked to generate the data reveals a significant flaw. The question asked was regarding your ‘favourable opinion of’ an individual and included people such as Oprah Winfrey. So in this case it was not a part to whole relationship (as the question was not ‘how do you favour this person over the others’,) and therefore, unlikely to be about presidential candidates (unless Oprah Winfrey is a standing candidate).

Figure 2



Careful consideration of how to present eClinical data is crucial to ensure decision makers and clinical teams can interpret data correctly to make the right assumptions. For example, a site may have a large number of queries in the EDC system. Does that mean it is a poor quality site? Not necessarily. The measure of the quality has to be a function of the number of subjects at that site. The site may have a large number of queries because they have a large number of subjects, each with multiple queries. A site with fewer queries but significantly fewer subjects will have a higher number of queries per subject. The number of queries is a measure of the work required, however, the number of queries per subject is a quality metric and has value in itself.

PhUSE 2011

Pre-attentive processing—tasks that can be done in 200 to 250 milliseconds with very little effort—adds value to data display as the eye quickly notices differences. The use of colour, grounding, closure, size and contrast can aid the interpretation of data, drawing attention to a particular area. These factors serve important substantive purposes beyond simply creating a metric that is visually attractive. However, design of dashboards often includes colourful banners and strong images detracting the user from the body of the information and overwhelming any impact the use of these tools may have had.

There are a number of pre-attentive attributes that can be used in data display. These include, but are not limited to:

- Colour
 - Hue
 - Intensity
- Position
 - 2D location
- Appearance
 - Orientation
 - Line length
 - Line width
 - Size
 - Shape
 - Added mark
 - Enclosure
- Movement

Careful selection and implementation of these attributes is advised to ensure the consumer is drawn to what is important. Colour or hue (what we traditionally think of as green, blue, red, yellow, etc.) and saturation of that hue (how intense that colour is from pale to intense) are important tools in designing visual displays; however, it is easy to use them inappropriately. To try and minimise the poor use of colour, careful consideration should be given to identifying colours that are acceptable to use for each situation. For example, if you know the audience for your data includes colour blind consumers, consider using black and white. Colour should not be used arbitrarily; a common mistake is a bar chart where each column is a different colour when there is no need to differentiate between the bars either because they are already labelled on the axis or are within the same informational category.

Charts convey a significant amount of data in a small space. They are not good for reporting absolute values, but instead, charts reveal the “shape” of the data allowing identification of trends, comparisons, exceptions, similarities, and differences between data. Charts and tables can be used together to complete a picture or example with tables showing current status of variables and charts showing how they have changed over time. If information is displayed using a chart, the user should ideally be provided with a mechanism by which they can access the underlying data.

Careful selection of chart types is necessary to ensure that stories hidden within data are appropriately revealed. In Figure 2 we saw an incorrect chart type selection where the data would have been more appropriately displayed using a bar chart. Understanding the nature of charts and the relationships they are best designed to portray is another fundamental principle in the design of information display.

The value of tables should not be overlooked when identifying the manner in which the data should be displayed. Not everything has to be displayed in a chart despite the current trends towards overly graphical dashboard displays. Tables should be used when the user persona is interested in ‘looking up’ individual values or comparing pairs of related values. Examples include number of subjects in treatment, number of sites activated, and number of CRFs incomplete. Tables should also be used when the user requires a level of precision that a graph cannot easily provide such as data to 2 decimal places.

There are a large number of considerations to be taken into account when designing information displays and, within the length of this paper, it is not possible to cover them all in great depth. For further information, please see the reading list provided at the end of this paper.

CONCLUSION

The consolidation of data and creation of new and powerful eClinical metrics outlined in this paper are not possible without a strong underlying technology infrastructure. Well-defined data architecture strategies, coupled with enabling integration/convergent technologies, are essential to ensure ease of access to data as well as a robust enterprise reporting tool to effectively display the information. Blending cutting-edge technologies with helping users achieve their business objectives through intelligent use of data and metrics is the key to a successful reporting strategy.

PhUSE 2011

It is essential that these key points remain at the forefront:

- To derive salient measures that truly inform decision making in the clinical trial process, it is essential to reevaluate and challenge common industry metrics.
- Providing information about trends in addition to current position is powerful in decision making because it not only shows how you got here but also where you may end up.
- One set of metrics does not fit all since different user personas have different needs. Although core similarities exist, understanding the differences between individual personas is even more critical.
- Consolidating data from multiple technologies provides value that cannot be obtained from individual databases.

RECOMMENDED READING

- Dave Zuckerman, *Pharmaceutical Metrics: Measuring and Improving R&D Performance*, (Gower Press, 1996).
- G. A. Miller, "The Magical Number Seven, Plus or Minus Two," *Psychological Review*, 63 81-97 (1956).
- Stephen Few, *Information Dashboard Design: The Effective Visual Communication of Data*, (O'Reilly, 2006).
- Stephen Few, "Data Visualization for Human Perception," (2010), [HTTP://WWW.INTERACTION-DESIGN.ORG/ENCYCLOPEDIA/DATA_VISUALIZATION_FOR_HUMAN_PERCEPTION.HTML](http://www.interaction-design.org/encyclopedia/data_visualization_for_human_perception.html)
- Perceptual Edge, [HTTP://WWW.PERCEPTUALEDGE.COM/EXAMPLES.PHP/](http://www.perceptualedge.com/examples.php/).
- N. Dowlman: Translating eClinical Data, ACT May 2011.
- Metrics Champion Consortium (MCC), www.metricschampion.org/default.aspx
- CMR International 2010 Pharmaceutical R&D Factbook.
- George A. Miller 1956 "The Magical Number Seven, Plus or Minus Two", *Psychological Review*.
- <http://www.turningleft.net/2009/11/fox-news-chicago-surveys-193-of-republicans-for-2012-presidential-race/>.
- Broad Street cholera outbreak, 1855 study by John Snow, http://en.wikipedia.org/wiki/1854_Broad_Street_cholera_outbreak.
- Charles Joseph Minard's 1869 flow chart on *Napoleon's Russian campaign of 1812*, http://en.wikipedia.org/wiki/Charles_Joseph_Minard.

CONTACT INFORMATION

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

Nikki Dowlman
Perceptive Informatics
Lady Bay House
Meadow Grove
Nottingham
NG2 3HF
UK
Work Phone: +44 (0) 115 844 0639
Email: Nikki.dowlman@perceptive.com
Web: www.perceptive.com

Brand and product names are trademarks of their respective companies.