Paper DV01

Visualizing clinical operational metrics to enable decision-making and accountability

Michael Gonzales, Biogen, Cambridge, US
Dan Boisvert, Biogen, Cambridge, US

ABSTRACT

Study milestone cycle time metrics are critical to understanding the health of a study, a department, and even a company. There are many ways to track, report, and visualize metrics, ranging from raw percentages to traffic lights. We developed a visualization scheme for fifteen key metrics, some of which are part of PHUSE's common Data Reporting Metrics[1].

In order to maximize the effectiveness of these visualizations, a dashboard was created using a popular tool, Tableau™. Equally as important, and as difficult, was the development of a process surrounding the use of this dashboard.

In this paper, we will detail the visualization scheme and the process while answering the following key questions:

- What impact does the dashboard and visualization choices have on the audience?
- What were the lessons learned?
- Where do we go from here?

INTRODUCTION

Our dashboard's primary purpose is to serve as a catalyst for study health discussion among leadership, study subject matter experts (SMEs), and other stakeholders. A meeting called the Operations Review Meeting (ORM) is held periodically throughout the year where this discussion takes place. The ORM dashboard uses Tableau and a custom web-application, called Pipeline, that captures study information, milestone planning, and actual dates.

The PHUSE Key Performance Indicators & Metrics Project idealizes what we are trying to achieve. A set of reporting metrics that are detailed but wide-ranging. Common enough that many life science companies can use but flexible enough to be tailored specifically for their purposes.

Tableau can connect to a plethora of data sources, has a highly active community, and an exceptionally simple interface for both developing and interacting with. It has limited capabilities in producing advanced statistical graphs, however, none were used for the ORM dashboard.

There are many visualization concepts and techniques, and perhaps the most well-known are described by Edward Tufte and his seminal work, *The Visual Display of Quantitative Information*[2]. The ORM dashboard incorporates many of these well-known techniques, diverges from others, and utilizes new ones.

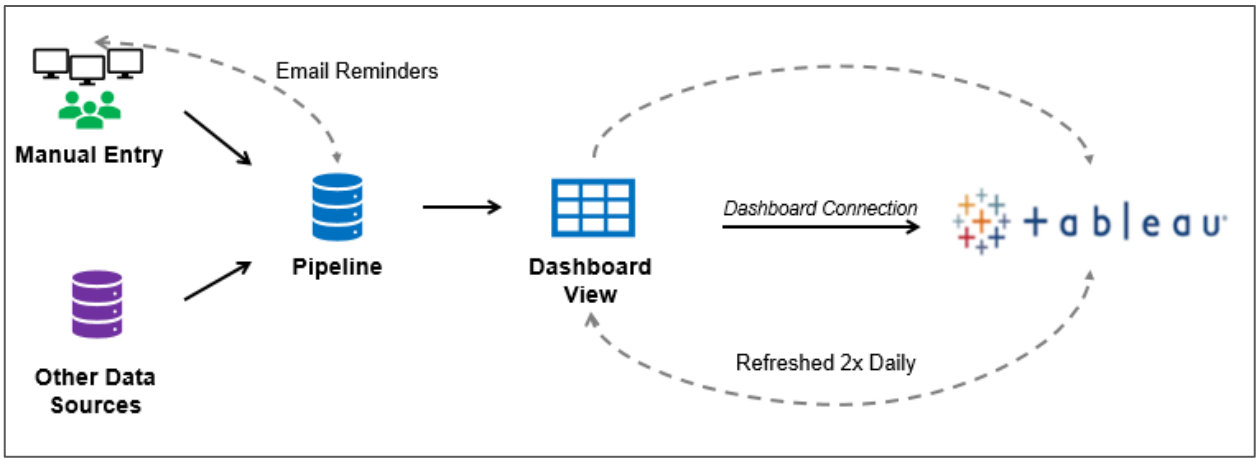
THE DASHBOARD

There are many critical milestones in a clinical study that help indicate a study's health. Fifteen metrics *(Appendix 1)* were prioritized to be visualized as they represent a broad overview of study deliverables. For the purposes of this paper, the last patient out to database lock (LPO-DBL) cycle-time will be used as the example. The rest of the metrics follow the same or similar visual schema.

TECHNICAL DETAILS

Tableau uses Pipeline as the primary data source through a MySQL connection. An extract of the data is automatically pulled twice a day to simulate real-time updates while allowing for some amount of consistency if the dashboard is viewed throughout the day.

Calculations of the metrics were originally performed using Tableau due to the ease of editing and lack of impact on performance. As the ORM dashboard has been used for more than a year, we recently created a normalized Pipeline view that contains all current calculations for the fifteen metrics as well as potential derivations (e.g. LPO to First Look, LPO to CSR, etc.). The dashboard now loads 500% faster and is more flexible. For example, all calculations are now pivoted, which allows for global filters. The normalized dataset provides more opportunities to be discussed at the end of the paper.



THE MAIN VISUAL

The picture below is showing one of fifteen metrics. However, the ORM Dashboard has six tabs:

- Guide that details how to read the graphs and what metrics are available
- A summary of all the metrics
- **Study Start-Up Metrics**
- **Study Maintenance Metrics**
- **Study Close-Out: Database Lock**
- **Study Close-Out: CSR Reporting**

Each of the bolded tabs have multiple metrics and each metric has multiple components:

a) %Met Bar Graph
b) Data Table
c) Average Line Graph
d) Box-Plot

OVERALL DESIGN

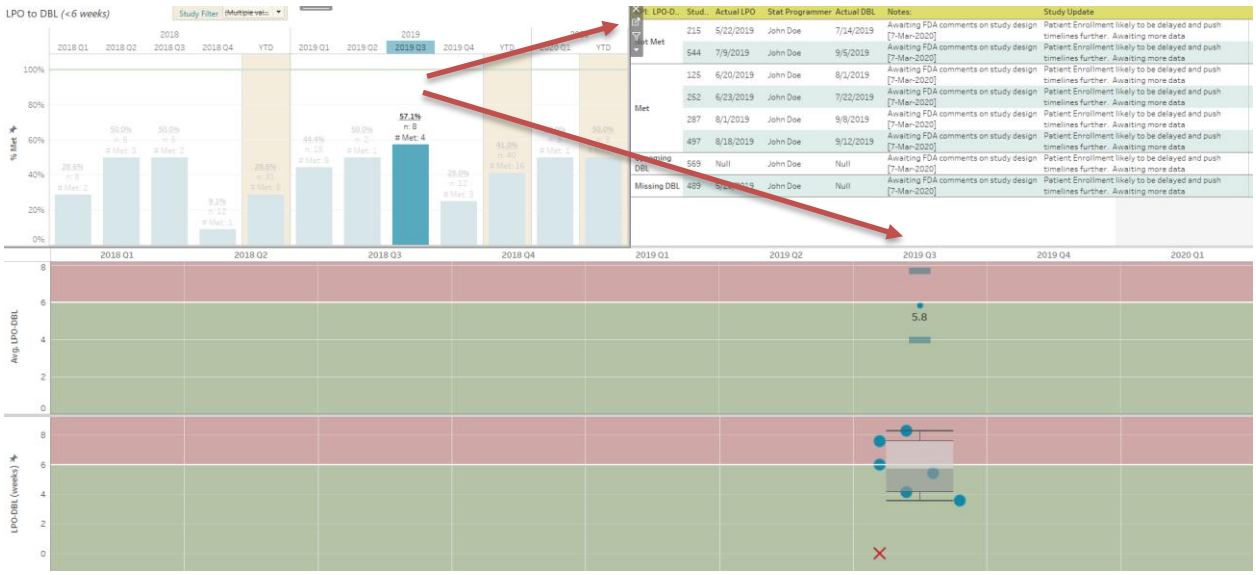
As one can see, none of the graphs are advanced from a statistical standpoint, due to the number of data points being quite limited as compared to typical clinical measures like number of patient visits. The intention of this design is two-fold: 1) to determine if we reached the metric threshold in a given quarter and 2) to generate discussion. Thus, a general aim to simplify the graphs, while making it as detailed *as needed,* was implemented through the:

1. *Layout:* Through a lifetime of reading and writing, many of us are naturally inclined to look at the top-left of a screen or document first[3]. We want the viewer to first go to the %Met graph, our most important measure. From there, the viewer has many options depending on need.

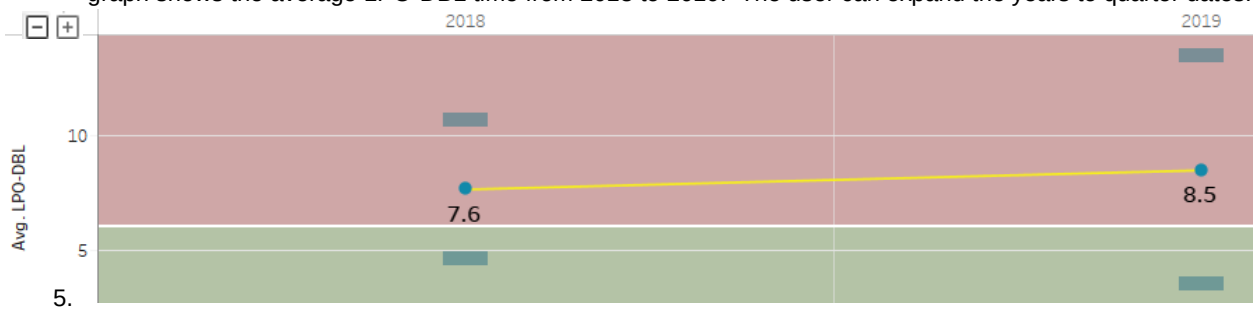
2. *Color*: Many guides will state that only two major colors, at most, should be used in a visual. One can look at the Amazon, Facebook, and Twitter logos to see that advice being widely followed (*95% of the top companies have logos with 1-2 colors*[3]). The visual remains appealing even if there are additional small variations of the two colors present (one could look at our Biogen logo (*Appendix 2*) as an example). There is a combination of teal, gold, and their variations. The green and red background colors are uniform, which our brains ignore if there is no focus on them.

3. *Filters:* Only one filter is visible to users; the study number filter. Although there are many ways to filter using the data (e.g. phase, indication, etc.), simplicity is crucial on focusing users.

4. *Interactivity-Filters:* One filter is visible but there are invisible filters that are intuitive to users. For example, if I want to see just the data table for 2019 Q3, then I simply click on the bar or on the date and the rest of the graphs will adjust accordingly. This removes the need for a date filter and encourages exploration of the visual. Individuals who actively discover something on their own retain the knowledge better than when passively instructed or trained[4].

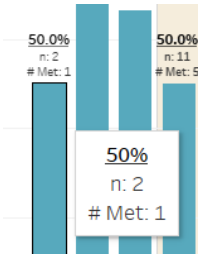


Another feature allows users to “collapse” time axes to summarize data. In the picture below, the average line graph shows the average LPO-DBL time from 2018 to 2019. The user can expand the years to quarter dates.



5.

6. *Tooltips:* A popular feature in modern dashboards are tooltips. Some tooltips offer critical information when space is needed, but they are also useful for taking screenshots or when presenting. In the example to the right, the tooltip is not displaying any additional information, but merely enlarges the existing label. This hovering over capability also encourages user exploration.

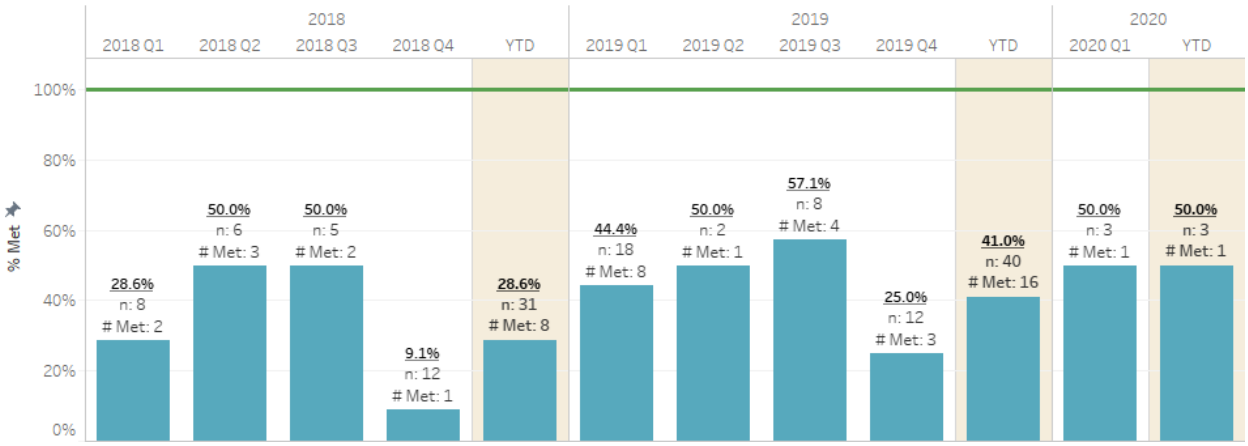


7. *Usability:* It goes without saying that potential users should be consulted on the creation of a dashboard. One useful tactic is to shadow a user as they are exploring the tool and trying to figure it out. It is important to remain silent and to not offer suggestions or instructions as to genuinely determine the tool's intuitivism.

8. *Location:* The ORM Dashboard is hosted on Biogen's Tableau Server and there is a direct link to it at our department's main Knowledge Management page. However, to maximize visibility, we also embedded the dashboard in our custom web application since that is where the source data is located. An association with the dashboard and Pipeline reinforces that idea that accurate study data results in accurate metric reporting.

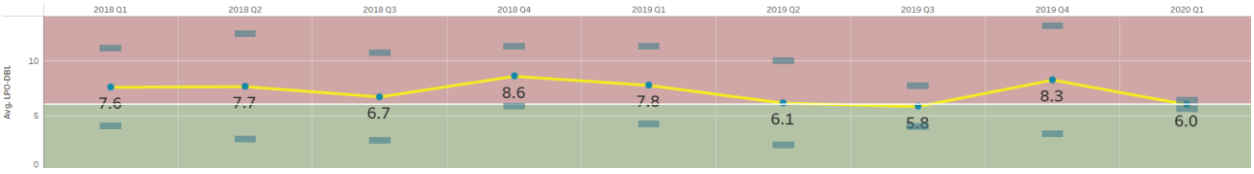
MET VS. NON-MET

This graph is an instant snapshot of the percentage of data points that met the threshold. In this case, it is the percentage of database locks which occurred less than six weeks after last patient out. The dates are generally anchored to the last date involved in the calculation. The green line being at 100% indicates that we expect all our database locks to reach this threshold. The labels indicate the number met and the overall number of studies with the relevant data points (in 2019 Q4, we had twelve database locks). If there were no database locks in a given quarter, then the label reads as "No Locks".



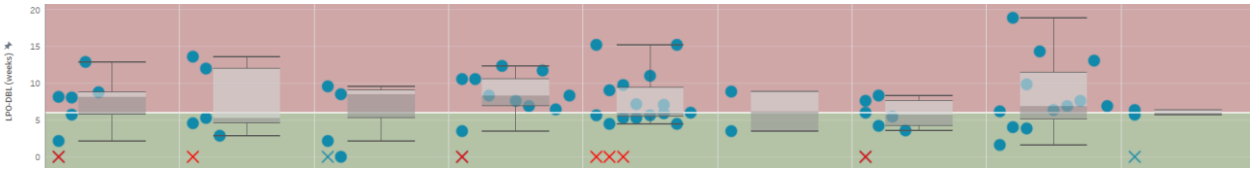
AVERAGE LINE GRAPH

Many stakeholders want to see a trending over time with their metrics. This graph shows the average, with the connecting line indicating to us if the average is going up or down. There are standard deviation bars (the light blue bars) to indicate the distribution of values, however, they are faded as they are not critical. The differentiation of the red and green backgrounds is the metric threshold, and in this case, six weeks. The graph very clearly shows whether the average is below or above the threshold.

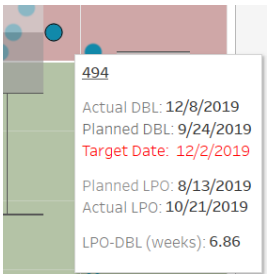


BOX-PLOTS

In addition to aggregated values, it is important to allow users to drill down with minimal work. Here we have a traditional box-plot with whiskers and medians. Each teal dot represents one data point (one study). The jittered dots allow the user to see how the average, median, and number of mets are being aggregated.



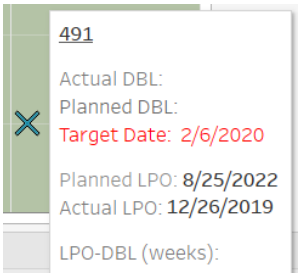
When the user hovers over a dot, a tooltip appears that provides additional detail. Not only can we see the relevant dates, but also the "target date", the deadline by which the metric would be met.



There are other nuances, even for simple metric calculations, that are worth highlighting. How do we manage missing data or late database locks?

The dates on the graph are anchored to the milestone date. If the milestone did not occur, then they are anchored to the *expected* milestone date. In the LPO-DBL example, the expected DBL date is six weeks after LPO (either planned or actual LPO). Such data points are presented as either a red X or blue X.

The red Xs indicate that the actual milestone has not occurred by the target date, and is blank in the database. We label those data points as "Missing". In the example to the right, we can see that the target date is six weeks after the LPO that has occurred. As of today, the actual DBL has not occurred or has not been entered in our database. Furthermore, the planned database lock is also blank, which can spark further discussion.



Similarly, the blue Xs indicate that we expect the milestone to occur but it is not yet due. In the example to the left, the target date of the database lock is six weeks after the actual LPO, in Q1 2020. Thus, the blue X is in Q1 2020.

Both the blue and red Xs can depend on planned dates, which themselves can be volatile. This can cause confusion if looking at the dashboard consistently, however, the benefits of transparency tend to outweigh the concerns.

TEXT TABLES

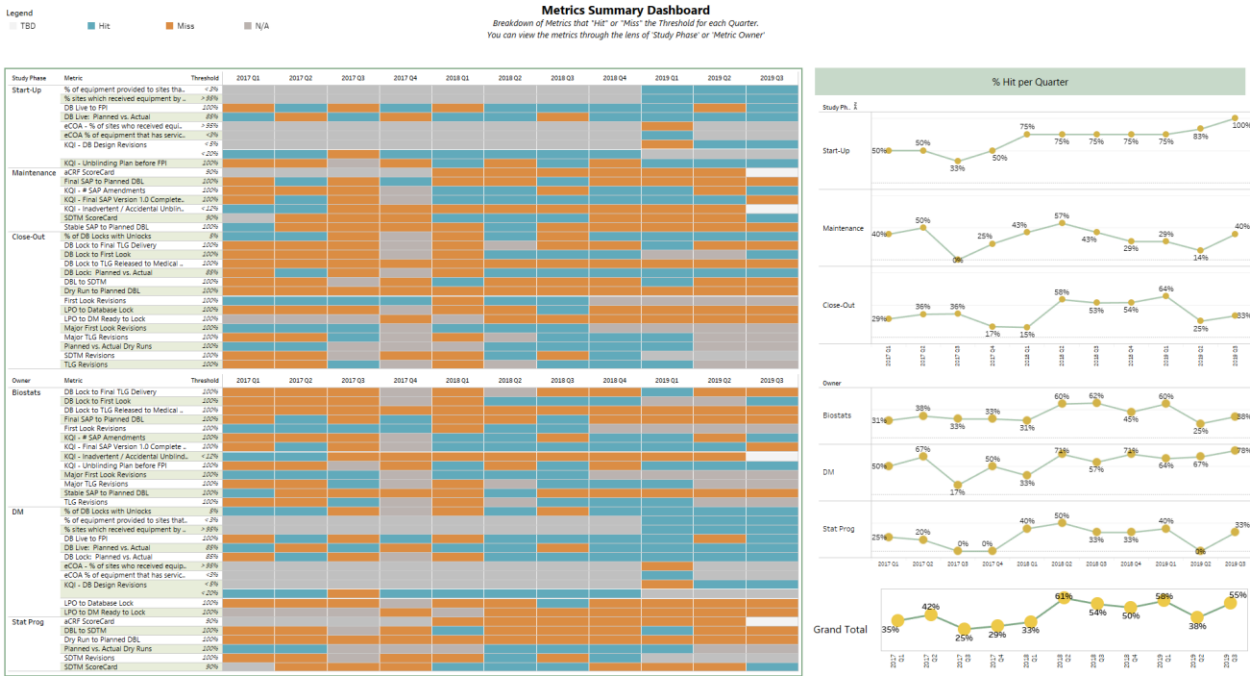
Although the power of visuals is universal, it is beneficial to provide a text option in case the user wants to quickly find a specific study for a specific metric. The table provides basic information and dates relevant to the metric. Study specific notes and comments are captured, sometimes illuminating why a deliverable will be late. The table is most useful when filtering through the other graphs.

KPI: LPO-D..	Stud..	Actual LPO	Stat Programmer	Actual DBL	Notes:	Study Update
Not Met						
	544	7/9/2019	John Doe	9/5/2019	Awaiting FDA comments on study design [7-Mar-2020]	Patient Enrollment likely to be delayed and push timelines further. Awaiting more data
	547	2/13/2018	John Doe	5/8/2018	Awaiting FDA comments on study design [7-Mar-2020]	Patient Enrollment likely to be delayed and push timelines further. Awaiting more data
	552	9/10/2018	John Doe	11/7/2018	Awaiting FDA comments on study design [7-Mar-2020]	Patient Enrollment likely to be delayed and push timelines further. Awaiting more data
	565	12/7/2017	John Doe	2/6/2018	Awaiting FDA comments on study design [7-Mar-2020]	Patient Enrollment likely to be delayed and push timelines further. Awaiting more data
	574	9/24/2018	John Doe	1/8/2019	Awaiting FDA comments on study design [7-Mar-2020]	Patient Enrollment likely to be delayed and push timelines further. Awaiting more data
	006	5/20/2018	John Doe	6/21/2018	Awaiting FDA comments on study design [7-Mar-2020]	Patient Enrollment likely to be delayed and push timelines further. Awaiting more data
	012	2/21/2018	John Doe	3/8/2018	Awaiting FDA comments on study design [7-Mar-2020]	Patient Enrollment likely to be delayed and push timelines further. Awaiting more data
	023	6/26/2018	John Doe	7/11/2018	Awaiting FDA comments on study design [7-Mar-2020]	Patient Enrollment likely to be delayed and push timelines further. Awaiting more data
	031	4/22/2018	John Doe	5/29/2018	Awaiting FDA comments on study design [7-Mar-2020]	Patient Enrollment likely to be delayed and push timelines further. Awaiting more data
Met	035	2/7/2019	John Doe	3/18/2019	Awaiting FDA comments on study design [7-Mar-2020]	Patient Enrollment likely to be delayed and push timelines further. Awaiting more data

SECONDARY VISUALS

There are many ways to aggregate the metrics themselves. We chose to do two ways: by study phase (start-up, maintenance, and close-out) and by functional ownership (Biostatistics, Statistical Programming, and Clinical Data Management). The visual is broken into two main parts: a heat map and a line graph of % of metrics that met their threshold over time.

The purpose of a heat map is to draw attention to specific areas. Which functions are having the most difficulty with meeting their metric thresholds? Which study phases? The line graph also contains a summary of all metrics and is meant to represent the health of our entire department over time, however, they are not definitive. Does a low percentage mean the study health is poor in a particular area? In reality, not necessarily. However, a striking visual and maximum transparency can help jump-start a discussion among different levels in the organization.



THE PROCESS

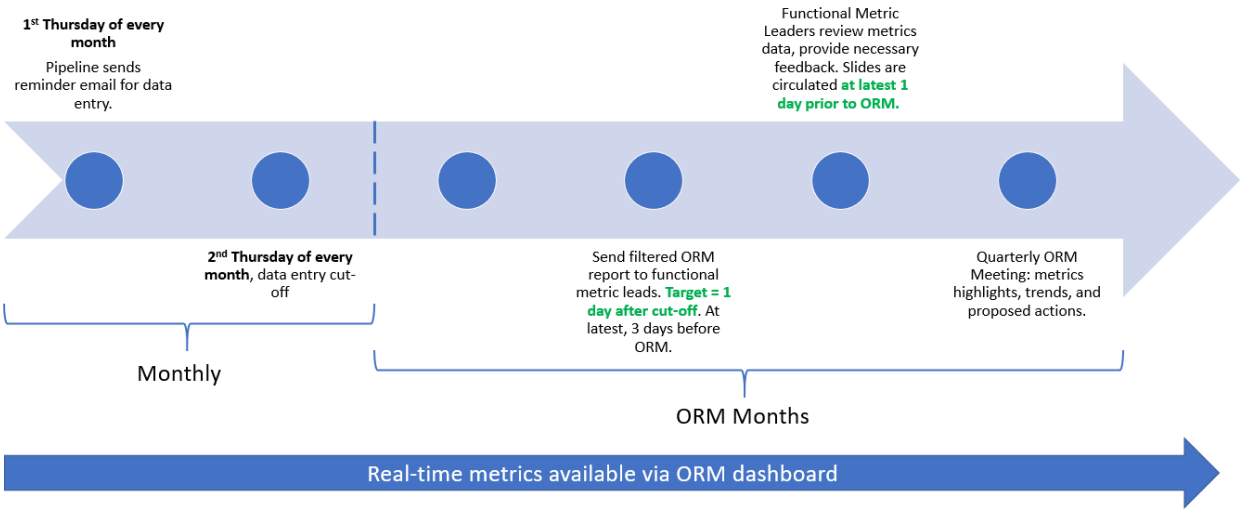
Metrics are challenging to present and discuss since they are time-dependent, owned by different functions, and reporting methods can vary wildly. How can we discuss metrics honestly and openly if we can't even agree on how we are going to measure and show them?

DEVELOPING THE METRICS

There are hundreds of different study metrics one can measure. We decided on fifteen as the most critical ones to report on. Each function is responsible for additional metrics, but a clear line of sight was needed from our leaders and a small amount is recommended for focus.

PREPARING THE DATA

The visuals are only as good as the data itself. Study teams are reminded through both manual and automated mechanisms to input study data and milestones. A schematic is included below.



OPERATIONS REVIEW MEETING

The ORM occurs quarterly and is attended by: biostatistics, data management, and statistical programming functional heads, process owners, and other leadership stakeholders.

Metrics are collated using the dashboard and shown during the meeting, accompanied by slides.

The dashboard reporting generates valuable discussion to identify the following:

- Which metric is underperforming and associated functional owner
- Owner determines plan of action
- Additional input from leadership, other functional owners
- Action item generated and followed-up on in the next quarterly meeting
- If action had impact metric in subsequent reporting

ADDITIONAL USES

The ORM dashboard was created for supporting the meeting and to have a central location for the most critical metrics. In the past year, however, other uses have been identified.

Missing Data – The red Xs let us know if a deliverable is late or if it hasn't been reported. We conduct a cleaning session monthly to fill in these potential gaps.

Ad-Hoc Reporting – The calculations can be applied to any metric. The threshold is the main variable but displaying metrics for other milestones in the same manner will increase visibility and familiarity with the visuals. An example is creating a stand-alone dashboard for Statistical Analysis Plan (SAP) metrics for the biostatisticians. Additionally, generating end of year metrics to celebrate (in forums like Town Halls, etc.) or highlight is very easy and quick.

CHALLENGES

There remain many challenges in using the dashboard from both a technical and procedural perspective.

TECHNICAL / VISUAL

Tableau was chosen for a variety of reasons, but mainly because the ORM was a pilot process, advanced programming was not needed, and how simple the design interface is. It is easy to show others how the dashboard is constructed, which leads to increased brainstorming and participating in the initial design, particularly with non-data scientists and leaders. This makes it especially useful to obtain buy-in. There are limitations, however.

Several difficulties were found when using Tableau to connect to the raw, unformatted data:

- Slow processing time and filtering from the number of calculations (over a hundred in all).
- Difficult to keep track of changes and the reasons for them.
- The milestone dates could not be pivoted easily, creating a lack of global date filters.

As the dashboard now connects with its own MySQL view, the processing time is much faster and all the metrics can now be grouped. However, challenges remain:

- How do we measure and visualize the impact of one or more milestones to another?
- How do we measure and visualize the impact of an action item or change in process to the metrics? Annotations are a possibility but are generally not flexible.
- Write-back capability is not possible in Tableau, yet. If the owner of the metric sees missing data, they must go to another system and correct it.
- Usage analytics are limited to number of views. More meaningful would be to analyze exactly how the users are interacting with the data.
- Use is limited to ORM. The dashboard is not user-friendly for specific study team members if they want to view a particular study.

It might be that other visual tools can manage these challenges and is the future visualization tool of the ORM dashboard. It is dependent on available expertise; however, the visuals remain viable no matter which tool is used.

PROCEDURAL

The ORM dashboard has increased transparency and accountability to the reporting of metrics. What has been discovered is that a more fluid metric threshold is needed to account for the multitude of reasons why a study milestone has appropriately been delayed. These types of “exception” data points can be detrimental to discussion. Annotations are a possible solution but are static and add to the overall manual process. One thing to be explored in future versions is to show the exception data point/outlier in the visual but not have it contribute to the metric average/median.

The next steps will center around truly capturing study health to foster meaningful solutions.

FUTURE OPPORTUNITIES

With a normalized dataset, increased reliability and familiarity of the visuals, and buy-in from leadership regarding effectiveness, there are significant potential opportunities to expand on how metrics are reported.

AUTOMATION

Since we can calculate expected “due dates” with metric thresholds, it would be a relatively simple manner to send notifications to the study team when a milestone is due. Ideally, the notification would be sent before the due date and the study team would have an opportunity to change the planned date accordingly. This would rely every milestone having a planned date.

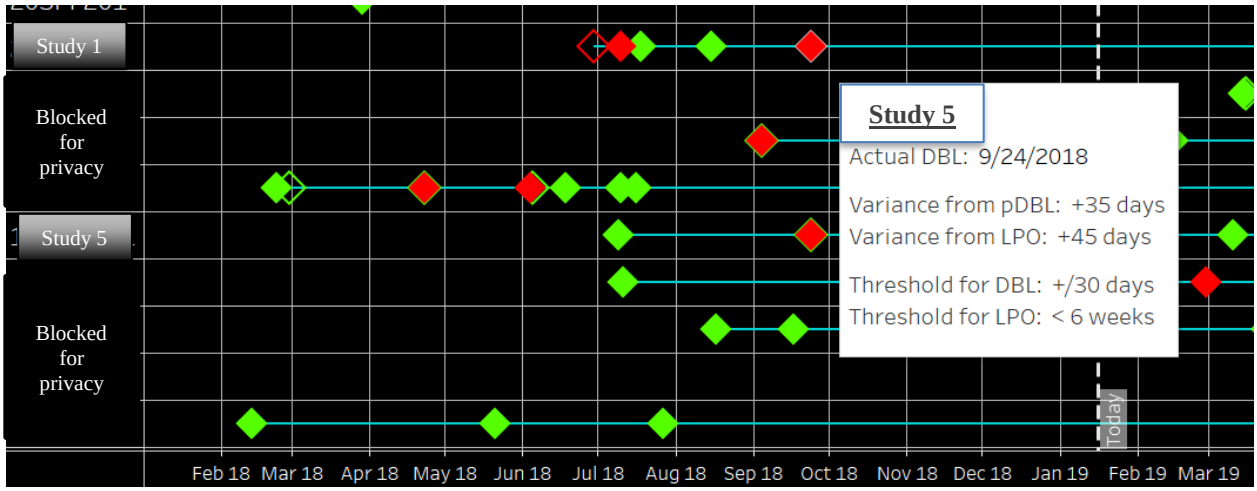
FUTURE TABLEAU ENHANCEMENTS

Visualization tools are ever evolving. One popular trend is voice command. Instead of clicking and scrolling through graphs, one can potentially use Tableau’s new Ask Data feature to “speak” to the dashboard and ask questions on demand. The graphs will change accordingly and is likely best suited for live presentations.

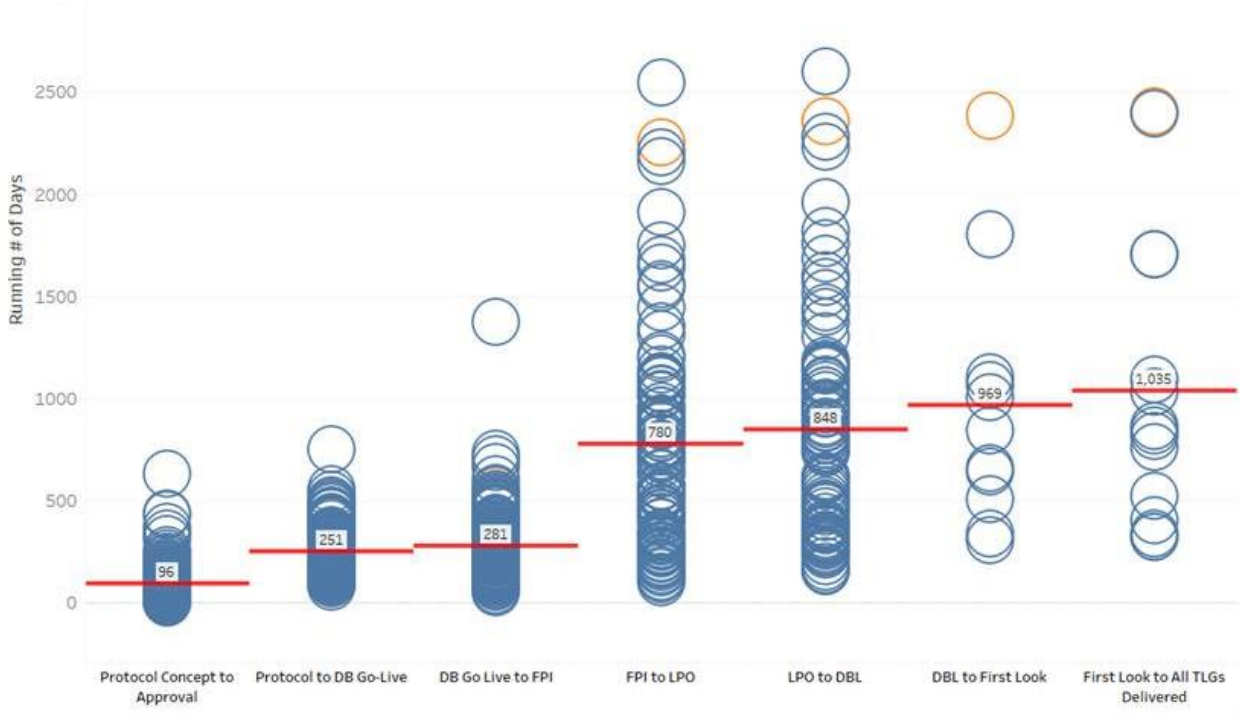
VISUALS & PROCESS

A normalized data set offers the ability to have one global anchor date for all milestones. Graphs do not have to be created separately for each metric, thus enabling the ability to create on-demand graphs. For example, the user can choose any milestone cycle-time they desire, even if it’s not part of the core fifteen metrics (for example, protocol approval to database go-live). The dashboard will then create the same graphs for those two variables on the fly. Every variable (like phase) would be available to slice and dice the data. Voice command and on-demand graph generation would go together in promoting use.

As mentioned prior, it is difficult to determine associations of metrics with one another. A step towards that direction is creating a timeline of study events. It would not measure the effect, but it could lead to discussion, especially between different functions. A proof of concept is included below. Study teams could see what has occurred and what will occur on an intuitive timeline visual. Green shapes are milestones that have hit their threshold (or will based on planned dates/unfilled shapes) and red shapes are milestones that are late or will be late.



Another proof of concept is adding all the times between milestones and lining them up (as seen below). Each circle represents a study and the red lines are averages of that particular metric. A user could select a study or another variable to see the corresponding cycle-times. Much work is needed to fine-tune which cycle-time should be employed as a default and how to operationalize these visuals.



Advanced analysis is worth exploring to detect correlations with all the operational study data. It would be laborious for a user to click through (or ask Tableau) all permutations in a scatter plot, although that could be a first step.

CONCLUSION

The greatest challenges in metric reporting are usually not solved by technology. They are people and process challenges that relate to accountability. Intuitive visuals have been a mainstay in advertising and marketing and have well-established practices. It will become increasingly vital to utilize these practices and develop new ones as we accumulate more and more data.

There is a place for visuals and dashboards to aid in helping people become accountable to the metrics. Objectivity, transparency, and appeal will draw users in, generate discussion, and take action. Action to improve our study health. Improve study health to help our patients, our ultimate purpose.

REFERENCES

1. PHUSE Common Data Reporting Metrics. https://www.phuse.eu/kpi-metrics-definitions-terminology-map.
2. Tufte, E. R. (2001). *The visual display of quantitative information.* Cheshire, Conn: Graphics Press.
3. Bouma, H., & Bouhuis, D. G. (1984). *Components of visual attention. Attention and Performance X.* Hillsdale, NJ.
4. Singh, Satyendra (2006). *"Impact of color on marketing".* Management Decision. 44 (6): 783-789.
5. Freeman, S. et al. (2014*). Active learning increases student performance in science, engineering, and mathematics.* Proceedings of the National Academy of Sciences, 111(23), 8410–8415.

ACKNOWLEDGEMENTS

We would like to thank Ritu Das (https://www.linkedin.com/in/rituparna-das-csm%C2%AE-b9b4b778/) for her help in improving the ORM Dashboard.

CONTACT INFORMATION

Comments, questions, and opportunities for networking are most welcome.

Mike Gonzales
Biogen
225 Binney Street
Cambridge, MA 02142
Mike.Gonzales@biogen.com
LinkedIn: https://www.linkedin.com/in/mikegonzales83/

APPENDICES

I. Top 15 Metrics.

Metric	Description	Threshold
Unblinding Plan before First Patient In (FPI)	All unblinding plans should be approved before FPI.	100%
Database Design Revisions	Percentage of studies that had databases built with revisions	5%
SAP v1.0 to Planned Database Lock (DBL)	• For Open Label Studies: before FPI (100%) • For studies with a "short" duration (<18 months FPI to LPO), 50% of duration after FPI • For studies with a "long" duration (>18 months FPI to LPO), 1 year before DBL	85%
SAP v1.0 before DBL	All SAP v1.0s to be approved before database lock	100%
Stable SAP to Planned DBL	• For Open Label Studies: before FPI (100%) • For studies with a "short" duration (<18 months FPI to LPO), 25% of duration after FPI • For studies with a "long" duration (>18 months FPI to LPO), 1 year before DBL	85%
DBL to SDTM	Duration of DBL to final SDTM	4 calendar days
DBL to First Look	Duration of DBL to First Look	10 working days
DBL to Final Table, Listings, Graphs (TLGs) Delivery	Duration of DBL to TLGs	4 calendar weeks
DBL to TLGs released to Medical Writing	Duration of DBL to Medical Writing	4 calendar weeks
Inadvertent Unblindings	blinded studies with >1 unblinding in the past 12 months	12%
Database Go Live before FPI	All databases should go live before FPI	100%
Database Go Live planned vs. actual	+/- 30 days from which database went live vs when it was planned to go live.	85%
DBL planned vs. actual	+/- 30 days from which databases were locked vs when it was planned to lock.	85%
Last Patient Out to DBL	Duration of LPO to DBL	6 calendar weeks
% of TMF documents uploaded within 35 days of finalization	Duration of document finalization/approval to Trial Master File upload (by function and overall)	80%

II. Biogen Logo