Paper DV09

Power BI in Clinical Data exploration, powerful visuals that answer relevant questions

DEBRUS Roxane, Terumo N.V., Leuven, Belgium

ABSTRACT

Nowadays, dynamic data visualization is vital for clinical study conduct and decision making. Power BI is a user-friendly business analytics service by Microsoft which provides interactive visualizations and business intelligence capabilities, offering unique integration options with Python, R, and SQL. Although a leader in BI and Analytics Platforms, its potential in clinical research is currently unexplored. To test it in this capacity and help our clinical and medical teams, we utilized it to design modern-looking, interactive dashboards, with the aim to improve the way we access, review and interact with data. As a result, we developed several tools that facilitate monitoring, analyzing and testing the study data in real time throughout the study, confirming that Power BI is perfectly suited to answer the needs of clinical and medical teams.

INTRODUCTION

Terumo Corporation was founded in 1921 in Japan. Since then Terumo developed more than 100 different medical devices in multiple fields. Currently, Terumo's clinical research is mostly in the fields of Interventional Cardiology, Interventional Oncology, and Peripheral Interventions.

In 2019, Terumo closed its biggest study ever, the e-Ultimaster Trial, which is one of the largest prospective worldwide registries in its field, which enrolled up to 37 000 patients. The device under investigation was a Drug Eluting Coronary Stent. It was an observational study, with a single arm, open label and 5 majors timepoint including baseline, procedure and 1 Y FU. The primary endpoint was to validate Efficacy and Safety based on a composite endpoint of different serious adverse event up to 1 year after the procedure.

eUltimaster's legacy data was captured through EDC system and extracted in SAS readable format. Over the years, all interim analyses were made on CDISC non-standardized data. However, before producing the final analysis for the clinical study report, it has been internally decided that an additional effort should be made in order to improve the traceability of the data and the transparency of the analysis results. If this topic interest you, please feel free to watch my other presentation shared with Silvia Faini from Liva Nova, about CDISC implementation in Medical Devices Clinical Research.

WHY DATA VISUALIZATION

After eUltimaster's database lock, Terumo organized a Publication Committee Meeting where the Top 15 enrollers were invited to discuss all possible future publications. During this meeting, we realized that the main investigators of the study had many questions, not all we could answer. In fact, they knew the status of their own patients, but were clueless of the status of the other sites.

The main questions were:

- How many site / countries participated to the study ?
- What was the contribution of each site / country to the study ?
- Number of subjects ? First/Last enrollers ? Inclusion duration ? Inclusion speed ?
- What is the quality of the data collected ? %LFU per site / per country ?
- What is the typical profile for inclusion ? Gender / Age distribution ? Lesion Characteristics ?

To preserve confidentiality of the data, we could not share the full database with the investigators. Nevertheless, we had to find a way to answer their questions.

In general, the most popular Data visualization platforms used in research and development are Spotfire from TIBCO and Qlik Sense. Before joining Terumo, I was working in a CRO where I used Tableau® to develop clinical dashboards. But Terumo was already using Power BI which is part of the Microsoft Office 365 suite. Data visualization activities had already been implemented in some departments at Terumo, mostly used in Finance and IT. I could have easy access to Power BI as it is available for Terumo employees. While outside of my comfort zone, I decided to investigate if Power BI could answer my needs.

WHY POWER BI

In 2020, Microsoft has been positioned as a leader in the Gartner Magic Quadrant for Analytics and Business Intelligence Platforms, and this for the 13th consecutive year. It is also the second year that Microsoft is considered the best in completeness of vision and the best in the ability to execute within the leaders' quadrant.

Figure 1 - *Gartner Magic Quadrant 2020 for Analytics and Business Intelligence Platforms.*

Power BI has many advantages including :

- User friendly interface
- Consolidate multiple data source (Excel, CSV, XML, Text, …)
- Interactive reports and geo-map visualizations
- Extended capacities by integration of R, Phyton and SQL
- Consistent reporting and analyses
- View reports across multiple platforms and devices
- In constant evolution

The only disadvantage met so far is that you cannot directly import your SAS datasets in PowerBI, so you first need to convert them in CSV or Excel file.

HOW TO DEVELOP DASHBOARDS IN POWER BI DESKTOP

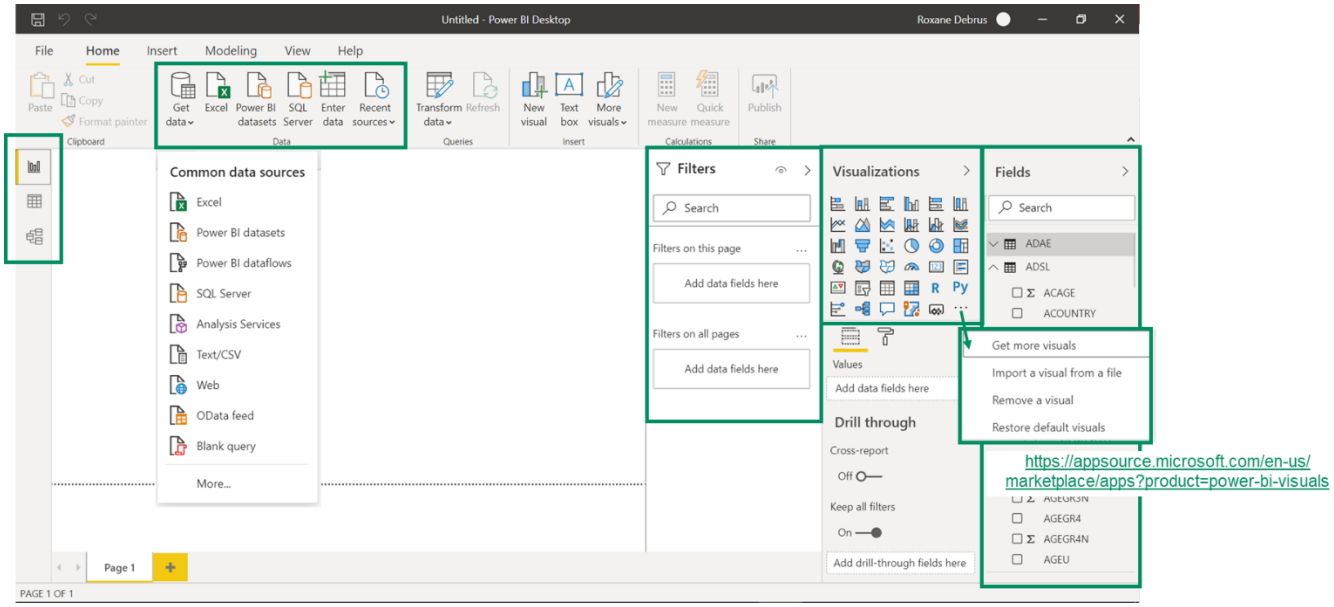


Figure 2 – *Welcome window of Power BI desktop.*

By using the data toolbox, you can import your data in Power BI. It can be an Excel file, or Text/CSV file but can also be imported from SQL server. It can even be a database from a previous Power BI dashboard. You can also enter some data manually if needed.

Once your data has been uploaded, the different datasets and variables appear on the right of your screen. As such it stays easily available. You can then choose the kind of visualization that you need. When you hover on each symbol; you get a description label to support your choice. The most used ones are bar chart, area chart, ribbon chart, scatter chart, pie or donut chart, maps, gauge, matrix table, etc. If you are looking for more visuals, you can click on the tree dots and select "more visuals". You will then be redirected to the appsource of Microsoft Market place where you can review many more visuals and download them (most are for free but not all of them). Once you selected the visualization, you can click it and it will be added in your workspace and you can easily drag and drop your variable in it. You can also predefine variables to be used as a filter on a specific visual, on the full page or on all pages (it can for example be population flags or known predictors).

On the left of the screen you have 3 possible view options, the first one is to see the dashboard, the second to see the database and the third one to see the relationship between the datasets.

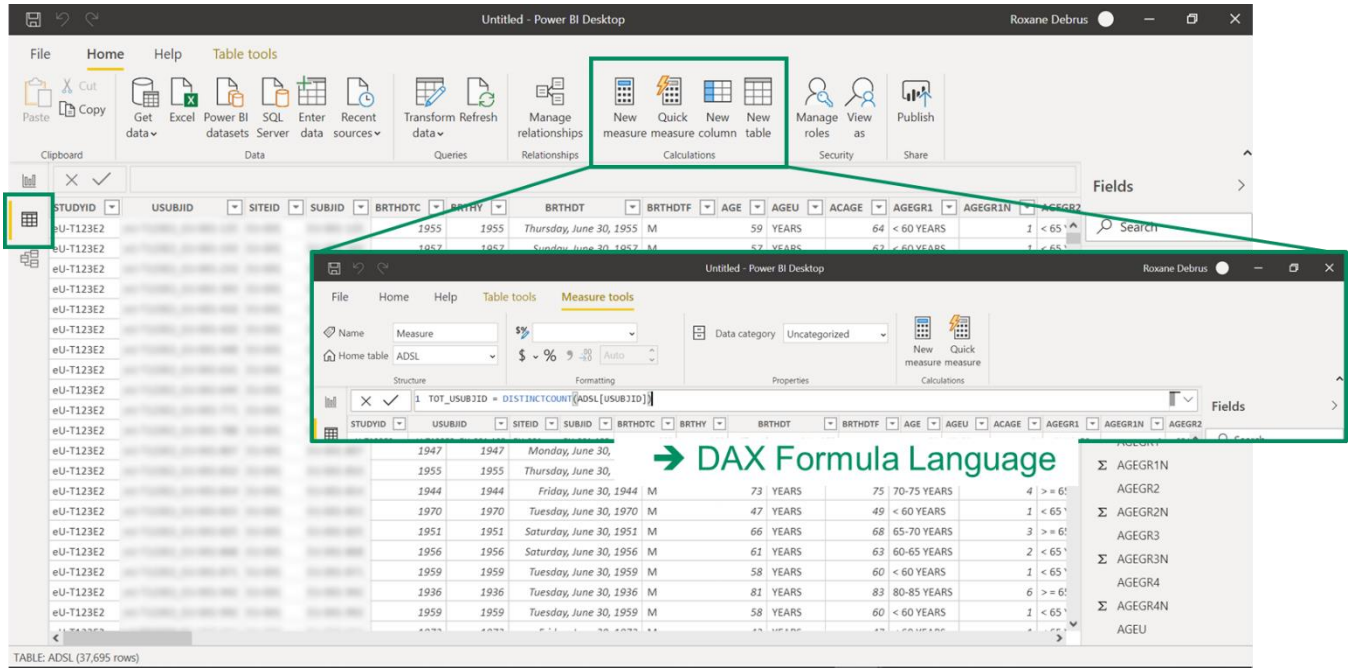


Figure 3 – *Database window of Power BI desktop.*

In the database view, you can have an overview of your data and filter like in every excel file by using the arrow. In the "Calculations" box, you can add new columns by deriving a static value or a new measure (by deriving a dynamic value, meaning that it will change when the selection will change), it can be for example total, mean, number of subjects, etc. Derivation algorithms need to be done by using Data Analysis Expression formula language, also known as DAX, which you might already have used while working with Excel Power Pivot.

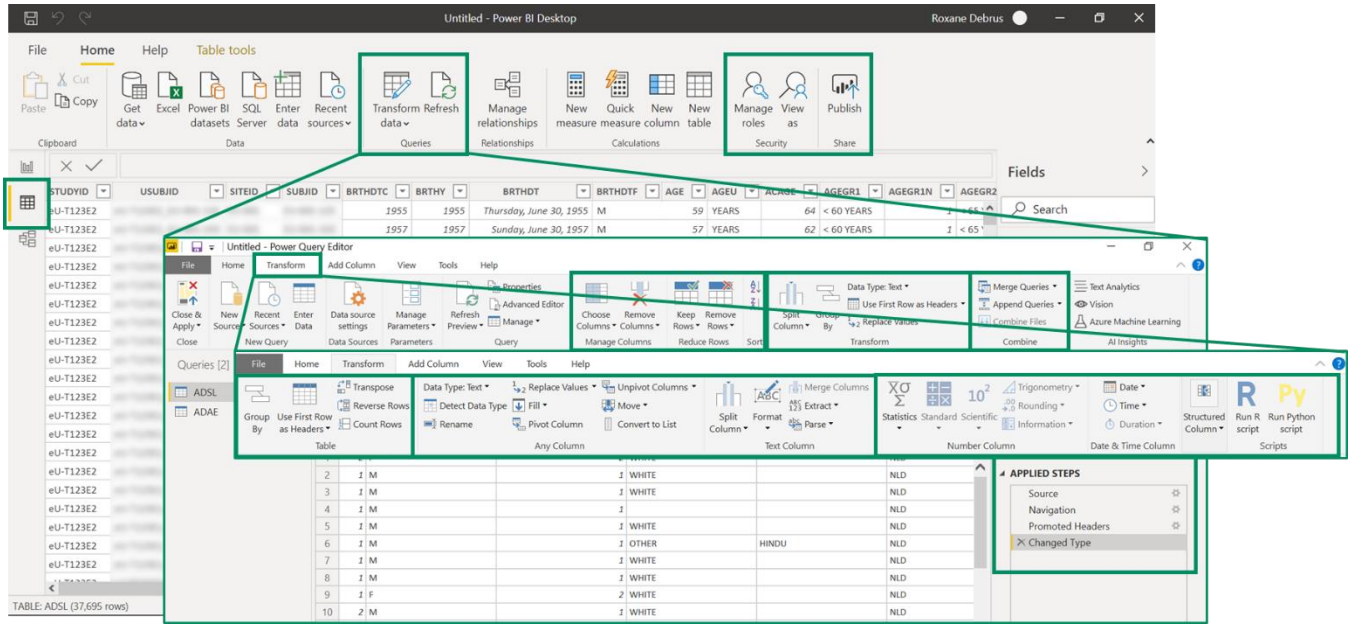


Figure 4 – *Database window of Power BI desktop, when transforming options are selected.*

In the "Queries" box, you can easily refresh your dashboard once your source data changed (new extraction) or you can define some transformation to be done. For each variable you get a description of variable type (character, numeric, date, etc.) but also a quick overview of the contained value (number of distinct and unique value). For example, the variable SEX has 2 distinct values and 0 unique. You can also see that it is not equally distributed and that one is more frequent than the other.

Transformation can be simple such as removing, splitting or merging columns, replacing values or changing data type. More complex transformation can also be made such as merge (left/right/full joint) or append datasets. By clicking on

the transform tab, even more options are available, such as transpose, group, count, rename, replace, unpivot, split, extract, parse and in the left of the screen you can see the option to add statistics, R or Python Script.
For each transformation made, Power BI defines a 'step' that you can modify or delete at any time. Once all the needed transformations are defined, when a refresh of source data is done all these transformations will be automatically re-executed.

Finally, in the "Security" box, you can assign different roles and as such define who can see which visual or data and then finally you can upload and share your dashboard with your readers.

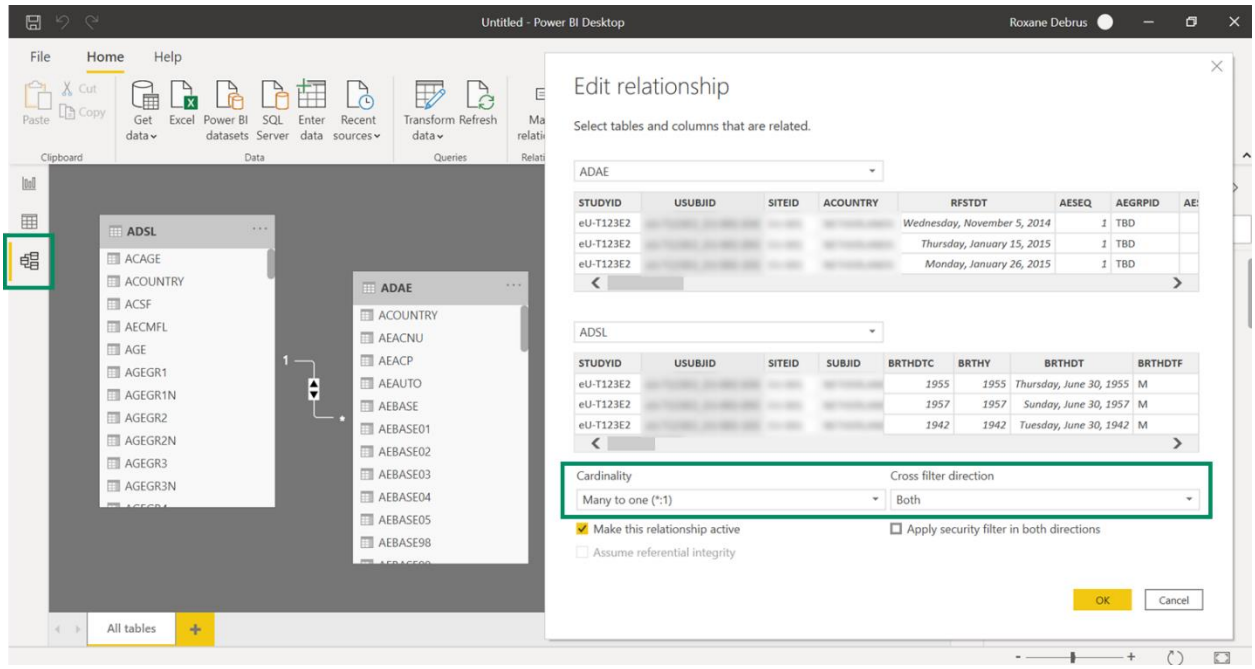


Figure 5 – *Relationship window of Power BI desktop.*

The last view is to define the relationship between the different datasets of your database. When editing relationships, you can define key variables to connect, the cardinality and the direction to be use for filtering. Defining proper relationship is very important to connect and ensure correct interactions in your dashboard.

TERUMO'S POWER BI DASHBOARDS

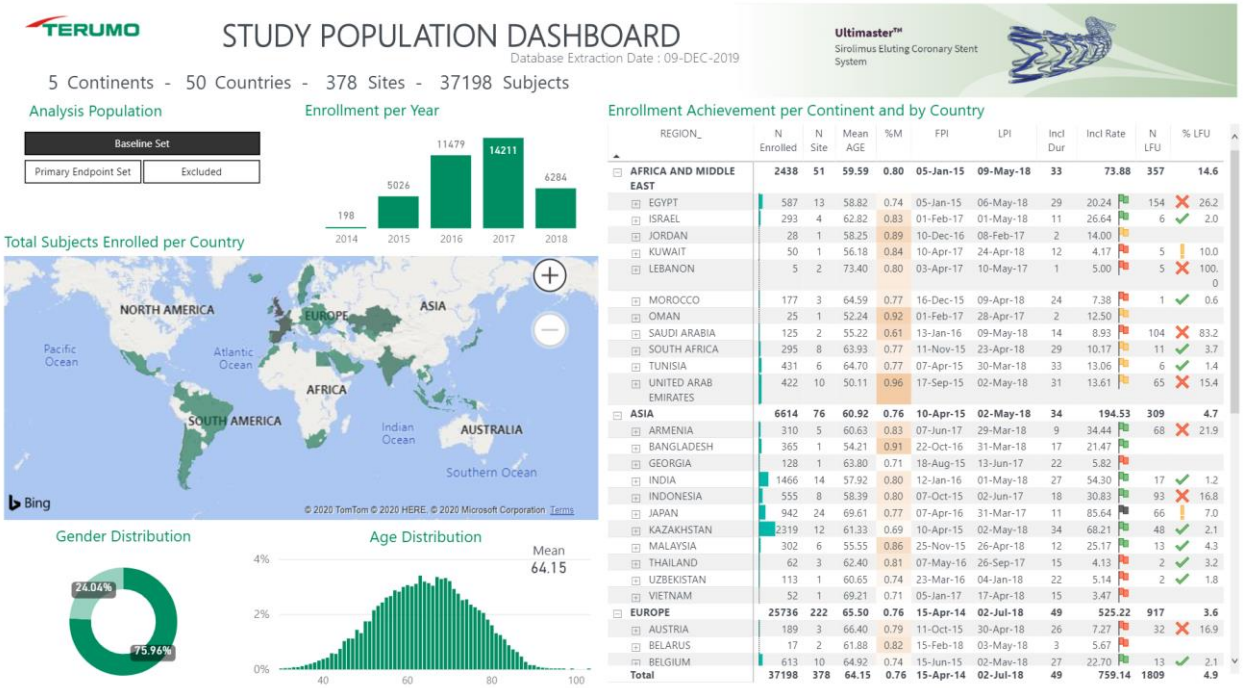
STUDY POPULATION DASHBOARD

The first project for data visualization was to try to answer all the Top15 enroller's questions with only one dashboard. Therefore, we created a Study Population Dashboard.

On the top, you can find number of continents, countries, sites and subjects. These are all dynamics variables. The bar chart gives you the enrollment per year, from the first subject enrolled in 2014 until the last one in 2018. The map represents the countries and the darker the color is, the higher the number of subjects enrolled. You can easily spot that in the top 5 countries with the highest numbers of subjects are UK, France, Netherlands and Spain. The fifth one being Kazakhstan.

The donut chart gives the gender distribution and the histogram the age distribution. In the table matrix, you can get additional details such as number of subjects enrolled, number of site, mean age and proportion of male, first patient in (FPI), last patient in (LPI), Inclusion duration (defined as number of months between first and last patient in), inclusion rate (defined as number of subject enrolled divided by inclusion duration), number of subject lost to follow up (LFU) and rate of lost to follow up (defined as number of lost to follow up divided by total number subject enrolled). All these are given per site (after clicking on the "+" sign), per country and per continent. In order to get these, you must generate a user defined hierarchy.

Some parameters have a conditional formatting to visually identify the highest and or lowest values. It can be a bar, a color in a cell background, or a colored mark such as a flag or symbol. Rows and columns can be sorted alphabetically or by ascending / descending order while reading the dashboard.



REGION_	N Enrolled	N Site	Mean AGE	%M	FPI	LPI	Incl Dur	Incl Rate	N LFU	% LFU
AFRICA AND MIDDLE EAST	2438	51	59.59	0.80	05-Jan-15	09-May-18	33	73.88	357	14.6
EGYPT	587	13	58.82	0.74	05-Jan-15	06-May-18	29	20.24	154	26.2
ISRAEL	293	4	62.82	0.83	01-Feb-17	01-May-18	11	26.64	6	2.0
JORDAN	28	1	58.25	0.89	10-Dec-16	08-Feb-17	2	14.00		
KUWAIT	50	1	56.18	0.84	10-Apr-17	24-Apr-18	12	4.17	5	10.0
LEBANON	5	2	73.40	0.80	03-Apr-17	10-May-17	1	5.00	5	100.0
MOROCCO	177	3	64.59	0.77	16-Dec-15	09-Apr-18	24	7.38	1	0.6
OMAN	25	1	52.24	0.92	01-Feb-17	28-Apr-17	2	12.50		
SAUDI ARABIA	125	2	55.22	0.61	13-Jan-16	09-May-18	14	8.93	104	83.2
SOUTH AFRICA	295	8	63.93	0.77	11-Nov-15	23-Apr-18	29	10.17	11	3.7
TUNISIA	431	6	64.70	0.77	07-Apr-15	30-Mar-18	33	13.06	6	1.4
UNITED ARAB EMIRATES	422	10	50.11	0.96	17-Sep-15	02-May-18	31	13.61	65	15.4
ASIA	6614	76	60.92	0.76	10-Apr-15	02-May-18	34	194.53	309	4.7
ARMENIA	310	5	60.63	0.83	07-Jun-17	29-Mar-18	9	34.44	68	21.9
BANGLADESH	365	1	54.21	0.91	22-Oct-16	31-Mar-18	17	21.47		
GEORGIA	128	1	63.80	0.71	18-Aug-15	13-Jun-17	22	5.82		
INDIA	1466	14	57.92	0.80	12-Jan-16	01-May-18	27	54.30	17	1.2
INDONESIA	555	8	58.39	0.80	07-Oct-15	02-Jun-17	18	30.83	93	16.8
JAPAN	942	24	69.61	0.77	07-Apr-16	31-Mar-17	11	85.64	66	7.0
KAZAKHSTAN	2319	12	61.33	0.69	10-Apr-15	02-May-18	34	68.21	48	2.1
MALAYSIA	302	6	55.55	0.86	25-Nov-15	26-Apr-18	12	25.17	13	4.3
THAILAND	62	3	62.40	0.81	07-May-16	26-Sep-17	15	4.13	2	3.2
UZBEKISTAN	113	1	60.65	0.74	23-Mar-16	04-Jan-18	22	5.14	2	1.8
VIETNAM	52	1	69.21	0.71	05-Jan-17	17-Apr-18	15	3.47		
EUROPE	25736	222	65.50	0.76	15-Apr-14	02-Jul-18	49	525.22	917	3.6
AUSTRIA	189	3	66.40	0.79	11-Oct-15	30-Apr-18	26	7.27	32	16.9
BELARUS	17	2	61.88	0.82	15-Feb-18	03-May-18	3	5.67		
BELGIUM	613	10	64.92	0.74	15-Jun-15	02-May-18	27	22.70	13	2.1
Total	37198	378	64.15	0.76	15-Apr-14	02-Jul-18	49	759.14	1809	4.9

Figure 6 – *Study Population Dashboard.*

When you hover with your mouse on a visual, additional information appears in the tooltip, like here, for % of grand total, FPI and LPI, and in the map, it gives the name of the country, the number of sites, and number of subjects enrolled. In the table matrix, you can also sort the rows and columns, For example, if you sort by descending mean age, you can see that in Asia, in Africa and Middle East Regions, the mean value is less than in the total population. If you filter on Asia by clicking on the row title, all visuals changes to only represent subjects enrolled in Asia, either by highlighting values compared to the total population such as the enrollment per year chart or fully filter such as the gender donut.

We can also select subjects enrolled in 2017 or change the analysis population. The condition formatting applied to the inclusion rate in the matrix table allows the reader to visualize it directly: black being the highest rate, then green, orange and finally red.

If you select United Arab Emirates, you can see that this country counts 10 sites, that most of their subjects were enrolled in 2017, that their proportion of male is higher than normal (96%) and their age distribution is shifted to the left, and so younger.

By reordering on inclusion duration, you can identify which countries had the longest inclusion duration and see if their rate of LFU are following the same trends or not. For example, you can see that Kazakhstan has a high inclusion rate and still a low rate of LFU.

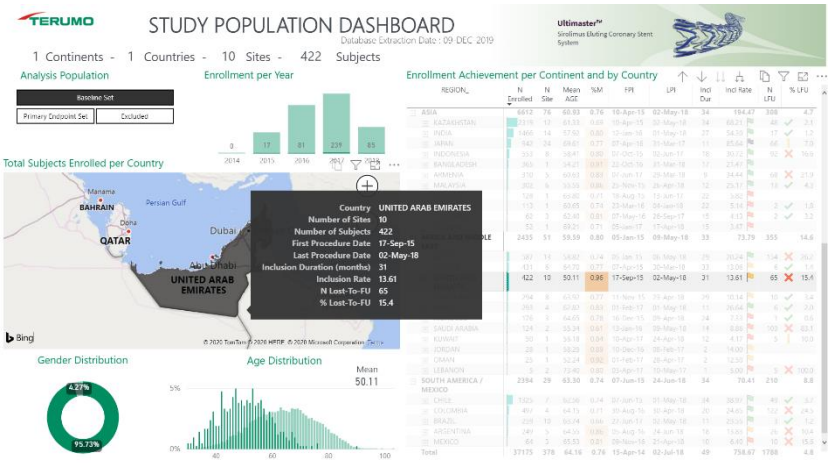


Figure 7 – *Study Population Dashboard updated when selecting United Arab Emirates country.*

Finally, you can make a multiple selection by pressing Control [CTRL] on your keyboard, in order to select different countries on the world map and the visualization will be updated accordingly.

SAFETY DASHBOARD

The second idea was to facilitate the monitoring and review of the Adverse Events of Specific Interest (AESI). Therefore, we built a “Safety Dashboard”. This dashboard had not been designed for eUltimaster as the study was already finished but for another study which is still ongoing. This study is about Roadsaver, which is a carotid artery stent. In this study, the adverse event of specific interest are death, stroke and revascularization.

As the occurrence of these adverse event (AE) are impacting the primary endpoint of the study, for all relevant adverse event reported in the eCRF, medical record and source documentation is requested to the site and provided to a clinical event committee (CEC) in order to adjudicate them.

Figure 8 – *CEC Process Illustration.*

Question from the study team were :

- How many Adverse Event have been adjudicated so far per site / per countries?
- Can we monitor the AESI rates ?
- Do we see difference in AESI rate based on baseline characteristics / predictors ?
- How different is the data reported in the eCRF by investigators from the Adjudicated Data by CEC members ? Related to Procedure / Related to Device ?
- Can we try to identify data inconsistency ?
- Under / Over reporting of AE ? Unexpected rate of a specific AE ?

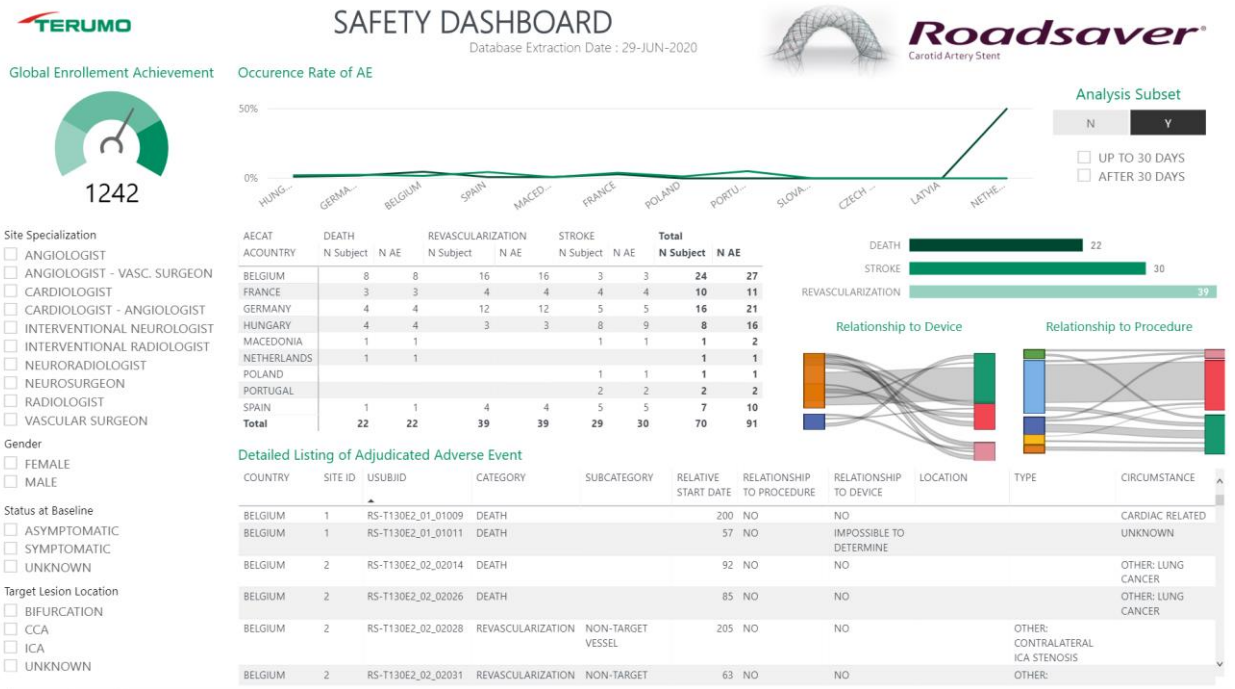


Figure 9 – *Safety Monitoring Dashboard .*

In the center of the dashboard is presented a table matrix presenting number of subjects with at least one AE and the number of AEs, per site, per country and in total. The upper graph gives the occurrence of death and stroke per country. Countries are sorted by decreasing number of subjects, so the reader can expect to see the abnormal numbers at the tail of the graph. At the time of the data extraction, there were only 2 subjects enrolled in The Netherland, and unfortunately one died, bringing the occurrence of death to 50%. This rate will decrease once the site will enroll more subjects.

In the bottom part is listed detailed information on all AE considered by the actual selection. In this study, the primary endpoint only considers AE occurring up to 30 days after the procedure, so a slicer has been added to allow the reader to focus more on these. In the bar chart, number of AEs are given per type of AE and specific rate are given in the tooltip. Predictors defined in the protocol have been used to define slicers located on the left side to filter on AE to be reviewed.

Sankey Charts have been included to compare relationships reported by the investigator in the eCRF (collected as “Causal Relationship”, “Probably”, “Possibly”, “Unlikely” and ‘Not related”) and the one assigned by the CEC members (as “Related”, “Not Related” and “Impossible to Determine”), one for relationship to Device and the second for relationship to procedure. This is an example of an additional visual available for free in the Microsoft Market.

LESION CHARACTERISTICS DASHBOARD

The third initiative tried to identify trends in lesion characteristics and how they could impact the outcome of the study.

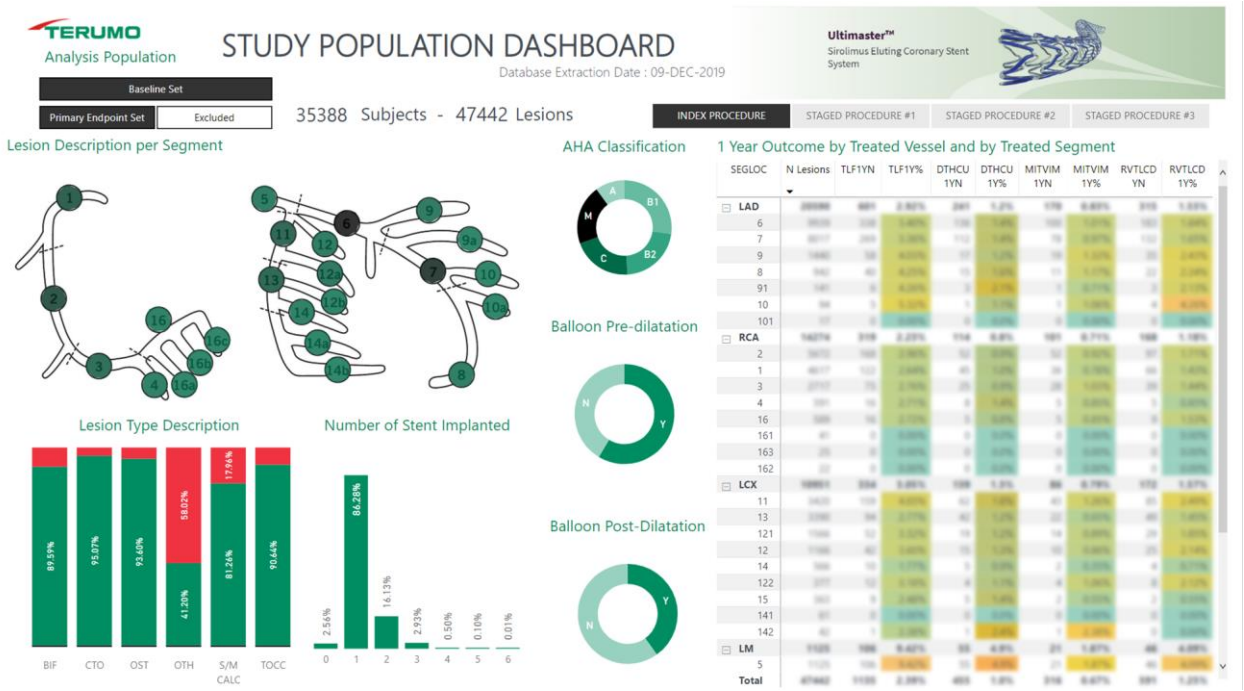


Figure 10 – *Lesion Characteristics Dashboard.*

As in the previous eUltimaster study dashboards, the counters and slicers are located on the top left part of the dashboard. The peculiarity of this dashboard is that it integrates a map completely drawn by the user. This custom shape map is present in the eCRF and the investigator must inform which segment(s) of which artery was(were) treated during the procedure. it is built in the same idea as the map of the world presented above. The darker the color, the greater the number of lesions treated at that location. You can easily observe that the top 3 treated segments are segment 6, segment 7 and segment 2. This custom map is also fully dynamic so when clicking on a segment, it will filter all the other values presented in the other visuals. When you hover on a segment, you can get the vessel name and number of lesions treated.

The first bar chart on the left is a 100% stacked bar chart representing lesion descriptions while the one on the right represent the number of stents implanted per lesion. The table matrix presents the number of lesions, number and rate of target lesion failure, cardiac death, MI related to target vessel and Clinically Driven Target lesion revascularization, which are the 3 sub-components of the target lesion failure. Numbers are given per segment and per artery, here again by using a custom hierarchy. As this data has not been publish yet, exact numbers can't be shared, however the color coding was kept allowing the interpretation, red being the higher numbers and green the lowest ones. Finally, additional parameters with potential impact on outcome are described in 3 donuts charts.

In general, you can see that we have around 10% of lesions including the bifurcation site. 58% of the lesion treated have a pre-dilation balloon while 40% have a post-dilation balloon. Lesion classified as B2 or C are known to be more difficult to treat and represent around 40% of the lesions treated. With a quick overview you can identify that the vessel with the highest occurrence rate of outcomes are both vein graft and left main. As vein graft (VGR) do not exist anatomically it is easily understandable that it is more fragile. However, for left main artery (LM), it is worth having a second look.

By clicking on LM segment, we can see where it is located and can better understand that as it is the artery that arises from the aorta and feeds all the blood to the left side of the heart, having a lesion there might impact much more than having a lesion further down the arch. We can also see that the proportion of bifurcation type lesions, increase from 10% to 45% but also the proportion of ostial and calcified lesions.

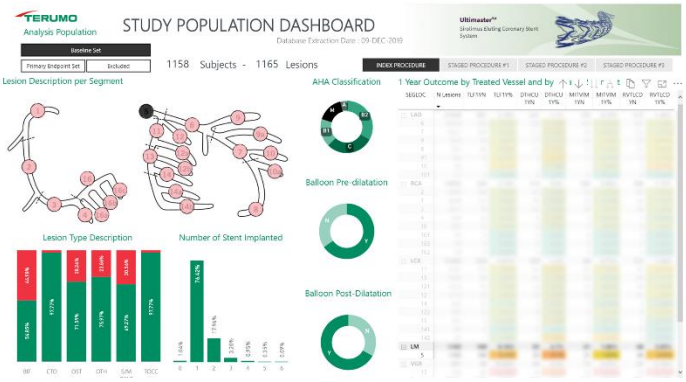


Figure 11 – *Lesion characteristics Dashboard updated when selecting Left Main Segment (LM).*

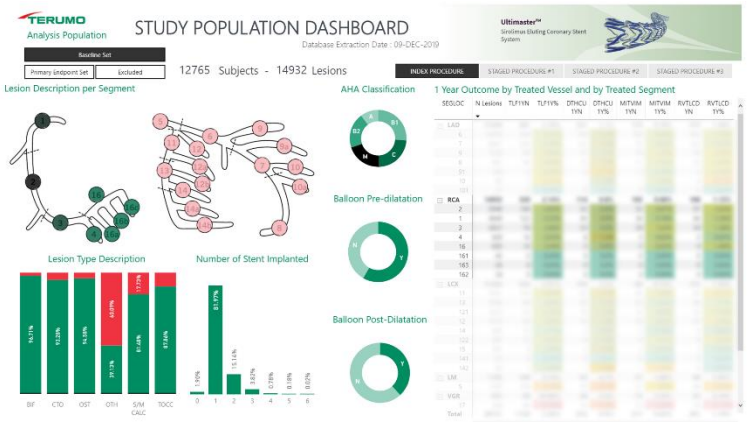


Figure 12 – *Lesion Characteristics Dashboard updated when selecting Right Coronary Artery (RCA).*

In the same way, when selecting the Right Coronary Artery (RCA), we can identify the segment to be considered and see that the proportion of bifurcation decreases to 3% while B2 or C lesion stay around 40%. If we select only B2 or C lesions, the highest number of lesions is in segment 6 and 7 and proportion of pre-dilation balloon uses changes from 58% to 69%. The outcomes rates of distal segments (i.e. segments 16a, 16b and 16c) are smaller than outcomes rates of proximal segments (i.e. segments 1, 2 and 3). This is also applicable for other artery such as LAD and partially to LCX.

COMPLEX PCI DASHBOARD

The last-born Terumo dashboard is the Complex PCI dashboard. For a subject to be considered as complex, 6 parameters have been identified: Bifurcation lesion treated with a 2-Stents technique (BST2), Chronic Total Occlusion (CTO), ≥ 3 Lesions Treated (LT3), Multi Vessel Treated (MVT), ≥ 3 Stents Implanted (ST3) and Total Stent Length ≥ 60 mm (TLG60). If any of the 6 is met, subject is considered as complex.

Questions from the medical team were first to understand how these criteria are distributed, and if they overlap. Then to assess their impact on specific outcomes and to identify which stent sizes are the most used.

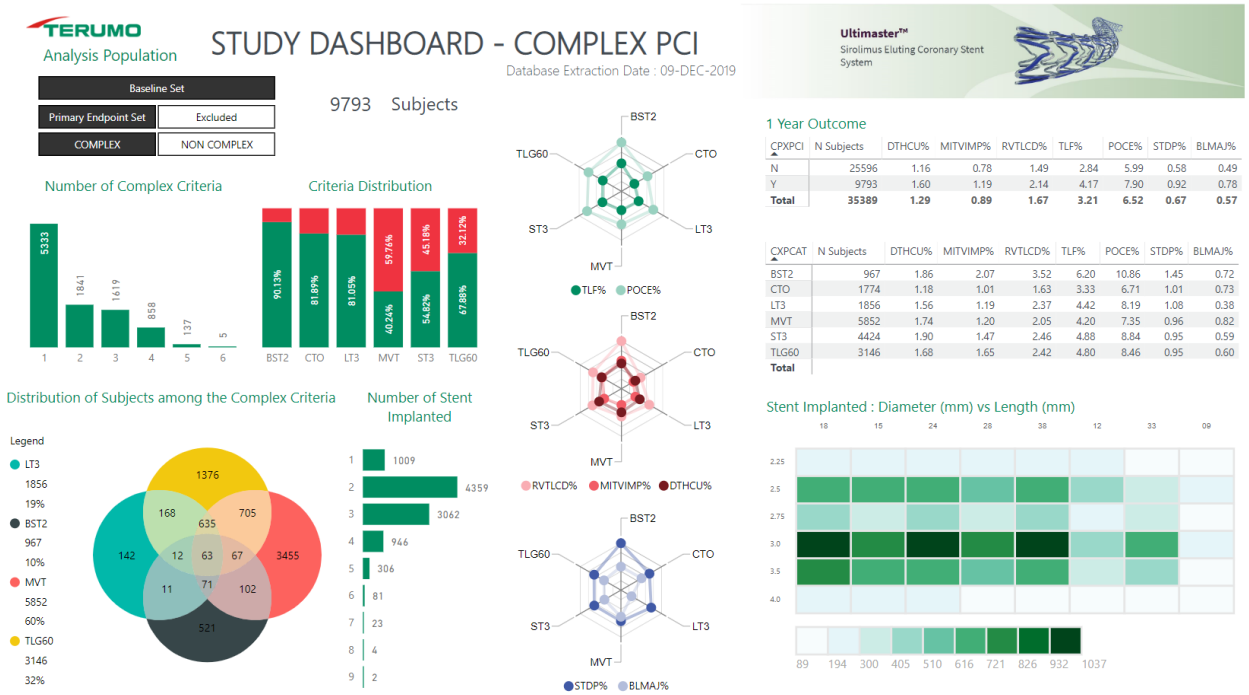


Figure 13 – *Complex Population Dashboard.*

In the eUltimaster study, approximately 1 on 4 subjects can be considered as complex. In the first bar chart, we can see that more than half of them have only 1 parameter, but we also see that 10% meet at least 4 parameters and 5 subjects have all 6 parameters. In the 100% stacked bar chart, we can see that the most frequent parameter met is Multi vessel treated and then more than 3 stents implanted.

In the Venn diagram we can observe how the parameters overlap. This visual is also available for free in the Microsoft Market, however its use is limited to 4 parameters and therefore a visual with the 6 parameters was not possible. The Table matrix give the number of subjects per parameter and the occurrence rate of relevant outcomes. In the middle, you can see 3 spider charts or radar charts. They represent the occurrence rate of each outcome per parameter. It is then easy to observe that target lesion failure is higher when subjects have a bifurcation lesion treated with 2 stent (BST2) and lower when the lesion has chronic total occlusion (CTO).

The last part of the dashboard represents the number of stents used to treat the lesion and the heat map identifies which stent size is the most used, presenting the diameter of the stent on the y axis and the length on the x axis. The darker the cell, the more frequent is it used. In this visual we can see that the 3.0 mm stent is the most used. This visual is also available for free in the Microsoft Market, however its use is limited to viewing the values in descending order. Therefore, the smallest length value appears on the right as it is rarely used.

To illustrate the complexity of the dashboard presented, let us look at the eUltimaster's relationship view. Here, you can see that ADSL is in the center of the picture and is related in a 1-to-1 or 1-to-many relationship to other dataset that can be themselves related to other dataset(s).

Here are some tips when defining relationships in database :

- To try to always keep a star schema, one subject level dataset in the middle, that can be used as a bridge to any other dataset.
- Avoid cross filtering direction with circular dependency, keep the connection flow as simple as possible
- Avoid using auto detected relationship otherwise you might end up with a big knot and lose control of the proper flow.

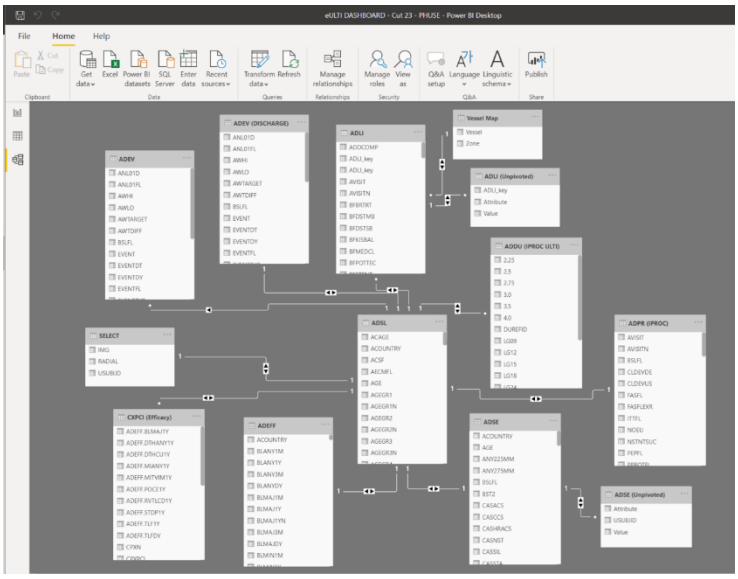


Figure 14 – *Relationship window of the eUltimaster project.*

NEXT STEPS

One of the strengths of Power BI is its extended capacities by integration of other language such as R. In this example, you can see how power bi reads an R script and creates the requested visual.

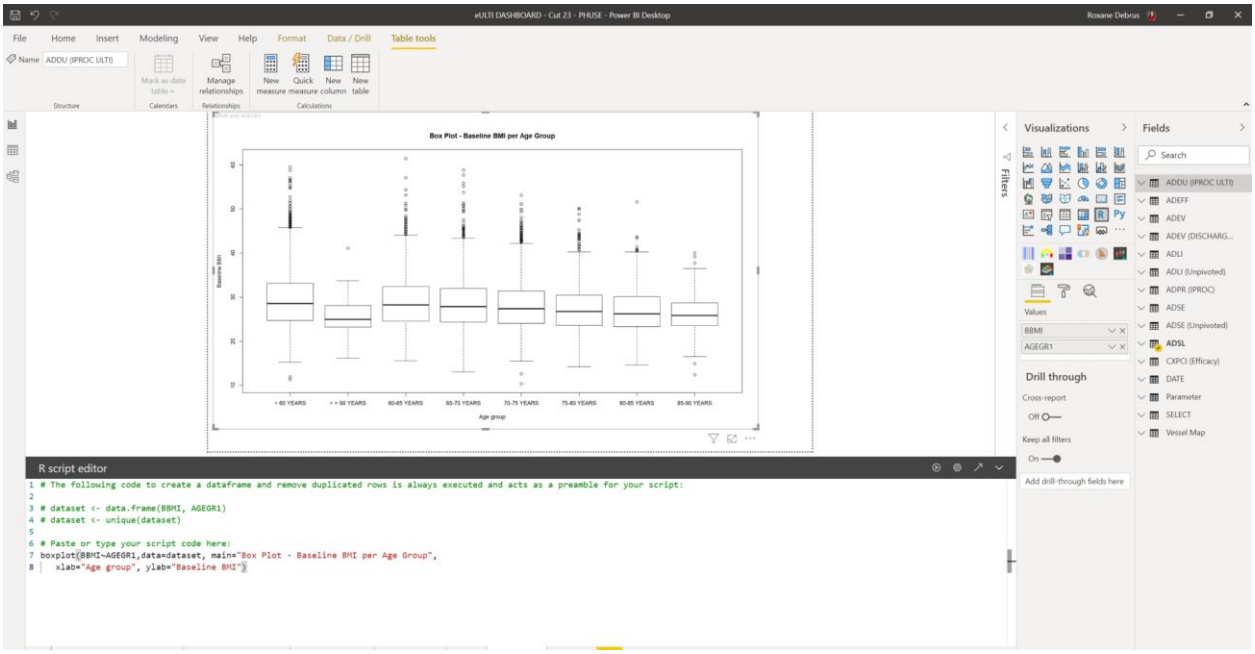


Figure 15 – *Integration of R script in Power BI Desktop in order to get user-defined visuals.*

Once we will have more developed R skills, we would like to push the limitations and program a Venn diagram, first with 4 categories, such as already presented in the Complex PCI dashboard, but then increasing complexity by adding a 5th category and then a 6th category.

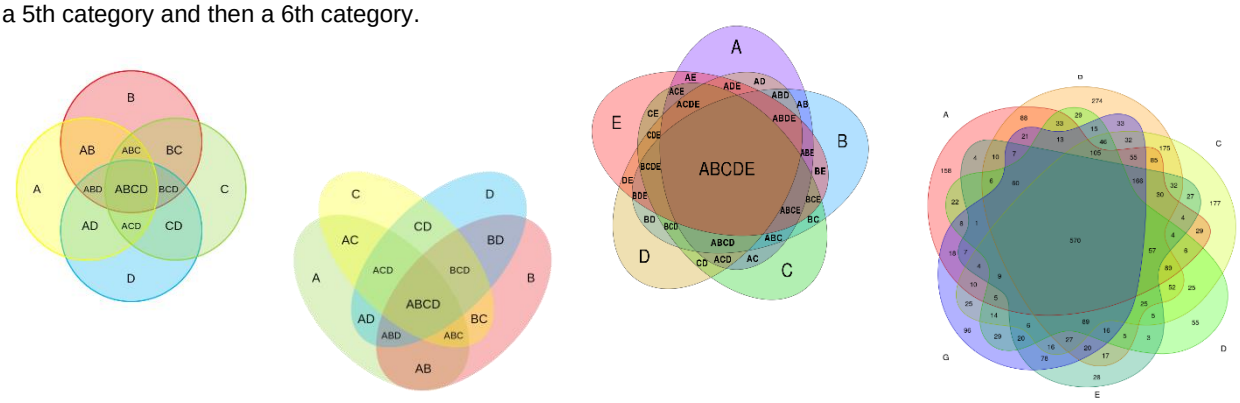


Figure 16 – *Venn diagrams with 4, 5 and 6 categories.*

With R script, we also would like to integrate some statistical output such as ANOVA test results or Kaplan Meier Curves.

Another great next step would be to use Power BI to automatically pool all studies performed with the same product. In fact, eUltimaster is the 5th study using this device since 2011, with different sample size, different follow up windows, different countries. By pooling different studies, we could get a more global picture. Of course, we need to consider that most of the picture will be taken over by the eUltimaster's study as the sample size was huge. However, this is just an example, and other devices than the eUltimaster stent have smaller studies where pooling can be very interesting.

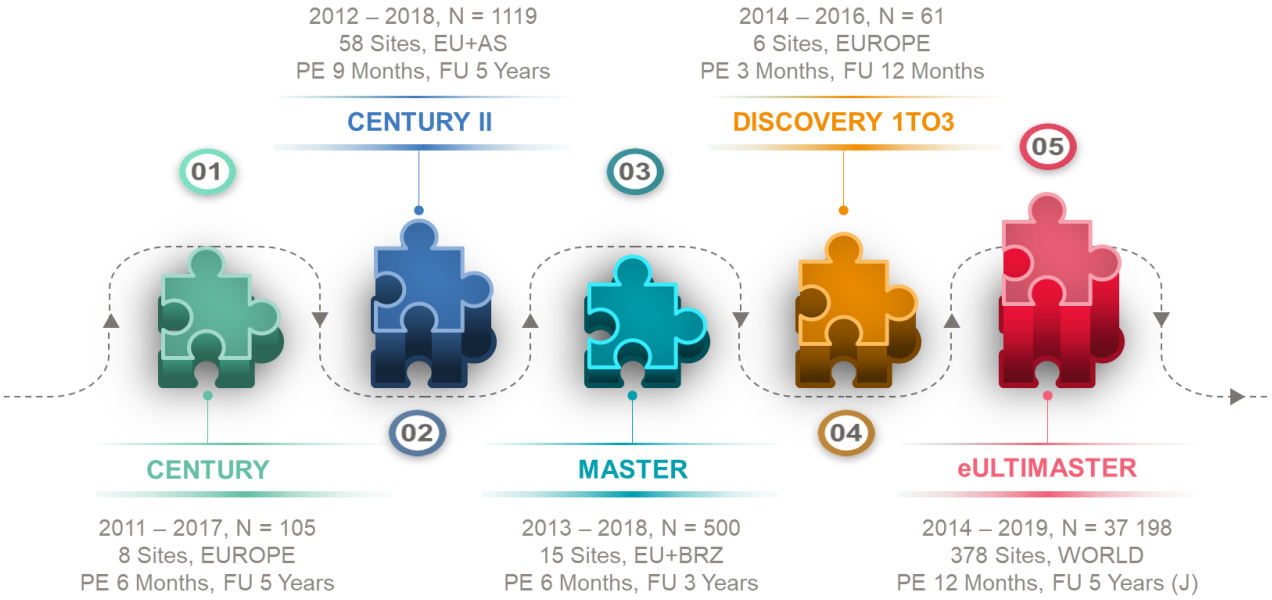


Figure 17 – *Previous studies made on the Ultimaster Stent from 2011 until today.*

CONCLUSION

As a conclusion, the most important is where you start. You need to clearly define your user's story, identify what is missing, what is the question you want to answer and what is your audience expecting. Keep in mind that for some of your readers, it will be the first time that they see such dashboards and graphs and you do not want to scare them.

Define the data to be used depending on your needs, if you want to refresh every day, maybe it is better to start on raw data. On the other hand, if your audience is composed of medical experts and you want to derive many parameters and avoid having to revalidate every step, maybe you should go for standardized data.

Start small and increase complexity progressively. Do not overcomplicate things. Do not link data randomly. Keep in mind that the level of complexity is not always proportional to the efficiency of your dashboard. Progress step by step and start from a medical question, does X impact Y?

REFERENCES

eUltimaster Terumo Page : https://www.terumo-europe.com/en-emea/clinicaldata/e-ultimaster-trial
Power BI website : https://powerbi.microsoft.com/en-us/
Create and use R visuals in Power BI : https://docs.microsoft.com/en-us/power-bi/visuals/service-r-visuals
2020 Gartner Magic Quadrant for Analytics and Business Intelligence Platforms : https://info.microsoft.com/ww-landing-2020-gartner-magic-quadrant-for-analytics-and-business-intelligence.html?LCID=EN-US&ls=Website

ACKNOWLEDGMENTS

I wish to thank all former and current colleagues who generously provided their support, insight and perspective that greatly helped to develop the presentation and improve the quality of this paper, but in particular: Bert Lefever, Lotte Jacobs, Frank Van Leeuwen, Tamara Sals, Natalie Bolton and Vladimir Borovicanin.

CONTACT INFORMATION

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

DEBRUS Roxane
TERUMO EUROPE NV.
Researchpark Haasrode 1520
Interleuvenlaan 40
3001 Leuven, Belgium
Work Phone: +32 16 38 12 81
Email: roxane.debrus@terumo-europe.com
Web: https://www.terumo-europe.com/en-EMEA