PhUSE EU Connect 2018

Paper DV09

Data visualization in a risk based approach to support decision making

Joke De Wever, Business & Decision Life Sciences, Brussels, Belgium
Sabine Jonas, Business & Decision Life Sciences, Brussels, Belgium
Irina Sargsyan, Business & Decision Life Sciences, Brussels, Belgium

ABSTRACT

Nowadays it is particularly important to have a visualization of data to answer the challenges faced by data management and clinical operations. It is part of a risk-based approach that associates cost to quality.
This paper will show how visualization on data queries can be included in the risk management process and support decision making.
Visualization enables easy identification of the sites that are less responsive or the type of queries resulting in poor resolution. As a result, sponsors may decide to set-up intermediate milestones for queries resolution, organize remote or on-site monitoring visits, provide additional training on a specific part of the study protocol or on the EDC system.
Moreover, using a standard structure (SDTM or sponsor-specific) for representing information about data queries allows pooling data from multiple studies using different Electronic Data Capture (EDC) Systems and helps sponsors adjusting their selection of participating sites.

INTRODUCTION

The ICH Addendum (R2) recommends using a risk-based approach to the quality management system. This includes but is not limited to the identification of critical processes and data, as well as the evaluation, control, communication, review and reporting of their related risks. The FDA has also issued a guidance for a risk based approach to monitoring.
Considering the large amount of complex data to be handled in clinical research, data visualization enables identification of trends and outliers. It provides the different stakeholders (data managers, medical monitors, clinical operations managers,…) with facts and figures to support decision making. Data visualization allows following-up study activities compliance and monitoring the Key Performance Indicators (KPIs) from study start to analysis.

Developing a dashboard on data queries provides an example on how visualization can significantly improve the efficiency of data management and data monitoring.
In this paper, we will present a query dashboard that has been designed with Tableau® but it can be developed with any other visualization software. This paper will not address the question about the choice of the visualization software but, of course, it will have an impact on the development process and on the resulting output. We will go through the development of dashboard, the interpretation of the graphs and show their potential benefits for study management.

DASHBOARD SPECIFICATIONS

The first step when creating a dashboard is to define the specifications. All aspects should be considered, from the development perspective to the end user's requirements. It is crucial to clearly identify the needs of the different stakeholders to meet the expected goals. It is also important to think ahead about the data structure to ensure efficient design and smooth reuse from one study to another. The features and limitations of the chosen visualization software must also be taken into account. It is strongly recommended to discuss and agree on the lay-out of the dashboard before starting its development. Formal User Requirements Specifications (URS) are recommended. The person in charge of writing the specifications should ideally combine an excellent knowledge of the data and good technical skills in the visualization software. This initial step is crucial for the success of the visualization.

DATA TO BE CONSIDERED

The data from the Electronic Data Capture (EDC) query audit trail must be used to identify all available variables in the system. Most of them are of interest and provide specific information that can be used (e.g. Date of creation, Category of Query, Status, Date of Closure,…). It is recommended to keep most of the source variables to enrich the dashboard. Indeed, a variable that is not relevant at this point in time could become useful for further projects. Moreover, it could be decided at a later stage to design other dashboards that would combine information from the query data source with data from any other topic of interest. As an example, a safety dashboard is often the first visual requested by Pharmaceutical Companies. It provides a good insight on (serious) adverse event status but can

also include details on the outstanding queries for a specific (serious) adverse event or a system organ class (SOC) if the query data is also available.
Displaying data from several studies could also be considered. If there are several studies done with the same sites, trends might be spotted by displaying data of all studies at once. As an example, information about sites responsiveness could be taken into account for site selection in further trials.
In conclusion, being as complete as possible in selecting the data to be included in the dashboard will help in further development decisions.

DATA FORMAT REQUIRED

Companies are using various EDC system which have their own specificities and nomenclature for queries related variables. To avoid re-work and loss of time, it is suggested to create a standard structure for those data (referred as XQ dataset). Most of the variables of the queries audit trail, especially the system variables, are common to all EDC system but are named differently. They need to be mapped to standard variables. This could be done by using an SDTM-like model (option recommended for CROs working with various pharmaceutical companies, see Figure 1) or a sponsor specific model (mainly used for visualization developed within a pharmaceutical company using one EDC system for all their trials). This mapping will then be part of the standard documentation (library) to be maintained by the company, for example when a new EDC vendor is contracted.

DOMAIN PREFIX	VARIABLE NAME	Label	FORMAT	CDISC CT	COMP. ALGORITHM	EXT. DICTIONARY	(CRF, eDT, Derived, Assigned)	CRF PAGE	CORE	COMMENTS	SOURCE DATA DOMAIN PREFIX	SOURCE DATA VARIABLE NAME	SOURCE DATA CODELIST	NOT MAPPED x	ONE TO ONE CONVERSION x or 99 or d	DURISOCONVERT x	DATE ISO CONVERT nD or nT	VARIABLES R n	CATEGORY RELATED n	FIXED VALUE	Additional programming comments
XQ	STUDYID	Study Identifier					Protocol, CRF		Req						99					STUDY001	
XQ	DOMAIN	Domain Abbreviation		DOMAIN			Assigned		Req						99					XQ	
XQ	USUBJID	Unique Subject Identifier					eDT		Req		QL				d						source field is blank; therefore populate with concatenation of XQ.STUDYID and QL.SUBJECT_ID e.g. study "STUDY001" and patient "1" becomes "STUDY001-000001"
XQ	XQSEQ	Sequence Number					Derived		Req	Sequence number (automatically generated) to ensure uniqueness within a dataset for a subject					d						is automatically generated
XQ	XQGRPID	Group ID					Assigned		Perm		QL	QUERY_NAME									
XQ	XQREFID	Reference ID					eDT		Perm		QL	QUERY_ID			x						
XQ	XQTERM	Reported Term					eDT		Req		QL	MESSAGE			x						
XQ	XQCAT	Category					eDT		Perm		QL	DESIGNATION			x						
XQ	XQSCAT	Subcategory					eDT		Perm												
XQ	XQSEV	Severity/Intensity		XQSEV			eDT		Perm						x						
XQ	XQOUT	Outcome		OUT			eDT		Perm		QL				d						SCAN(STATUS,1,' ')
XQ	VISITNUM						eDT		Perm		QL				d						SCAN(VISIT,-1,' ')
XQ	VISIT						eDT		Perm		QL	VISIT									
XQ	XQSTDTC	Start Date/Time of Observation	ISO 8601				eDT		Exp		QL	DATE_CREATED			x		1D				
XQ	XQENDTC	End Date/Time of Observation	ISO 8601				eDT		Exp		QL				d		2D				INTNX('DAY', DATE_CREATED, DAY_OPEN)
XQ	XQSTDY	Study Day of Start of Observation			XQ.XQSTDY		Derived		Perm						d						
XQ	XQENDY	Study Day of End of Observation			XQ.XQENDY		Derived		Perm						d						
XQ	XQDUR	Duration	ISO 8601				Derived		Perm		QL	DAY_OPEN									

Figure 1

EXPECTED USE

The first step in the dashboard development is to analyze and identify the end-users and the purpose of the dashboard. Trying to put yourself in the place of the end users will give you another perspective of the work to be achieved. As an example, the requests from a data manager, a clinical operation manager or a medical monitor will not always be the same, even if quite similar. Understanding the users' needs and customizing the dashboard to fit them is the key for success.
The ultimate objective of developing a query dashboard is to provide a tool for the study team to control the query resolution progress and identify any issues with data cleaning. These objectives can be divided in several goals which cover the monitoring of:

- Overall data cleaning status
- Country compliance
- Site compliance
- Data Managers activity
- Medical Monitoring activity
- Resolution of queries from a specific category (e.g. SAE, Medical Coding, Medical Review, Protocol Deviations)
- …

VISUALS AND THEIR LAYOUT

A good knowledge of the study protocol and study design is important to develop ad-hoc data visualization tools. Even for the query dashboard which is quite common to all types of projects, there are unlimited possibilities for the types of visual and layout.
Specificities from study design, study endpoints, clinical operations could help you defining the most suitable visuals for your queries dashboard. The complexity of the study (e.g. number of visits, number of treatments arms, treatment schedule), the safety profile of the drug (e.g. (S)AE reporting, potential specific monitoring such as cardiac tests,

tumor assessments, laboratory tests…) and the operational aspects of the study (e.g. number of participating countries/sites/subjects) are impacting the layout of the graphs. The end users can also request to highlight information reflecting problematic issues noticed with previous studies, participating countries/sites (e.g. sites that are responsive but only the very last days before a cleaning milestone, sites not willing to address queries without the help of the medical monitor).

All those aspects should be considered upfront to ensure that visuals are easily readable and not leading to misinterpretation.

A multitude of visuals are possible and each of us has his/her own preferences (e.g. pie chart versus treemap, gradient of colors versus color range). The development team should focus on the users' requirements to deliver a tool that meets the expectations and presents a design that is suitable for the majority of users. Consistency across dashboards' layout should also be maintained when delivering for the same client.

MAINTENANCE AND DATA GOVERNANCE

The maintenance process of data visualization tools should be part of the company data governance process. Roles and responsibilities, as well as changes control, should be clearly defined.

The ideal approach is to have a library of general dashboards that are quite standard. Those ones will be maintained over time when any required update is identified. They will then serve as basis for the creation of more specific dashboards (study specific with additional client expectations related to study phase, trial indication, study endpoints,…)

A query dashboard is a standard dashboard. However some adjustments might be required according to the EDC system used or even to new EDC version release. This will be mainly done by amending the mapping of the EDC variables and sometimes also by customizing the layout of graphs.

GLOBAL QUERY STATUS

Who hasn't been facing questions such as:

- Which are the visits or eCRF forms that are resulting in most of the queries?
- Which are the most problematic sites?
- Which country/site is impacting the most the cleaning status?
- What is the average duration of query resolution by site?
- What is the average duration of query closure by data management?
- When was the oldest open query issued?
- How many queries are pending action from medical monitor?

The list is not exhaustive. Having a strong data visualization tool that can answer those questions by a simple look at the dashboard helps managing the study more efficiently. All users can find by themselves the information they are looking for and immediately obtain a real-time status.

Figure 2 (note: this dashboard gives a snapshot of the data in October 2017)

The first objective of the query dashboard is to provide the study team with real-time status of the data cleaning process (see Figure 2) and the number of open queries (see Figure 3). For that purpose, the query dashboard is including KPIs. One of them is usually the percentage of open queries (see Figure 4). This information is particularly important for the preparation of interim and final analysis for which both medical monitors and data managers are collaborating closely to have all data clean.

Another KPI could be the percentage of queries from pre-defined categories (e.g. Serious Adverse event, Protocol Deviations, Medical Review,…) (see Figure 5). Those metrics are indicative of the different activities within the cleaning process and provide the team information about any delay in the process. However, they are also correlated with the complexity and safety profile of the study. Indeed the extent of eligibility criteria could be resulting in more protocol deviations queries. On the other hand, an investigational drug that is known to have a high toxicity profile will show more (serious) adverse events. The study will then require an intensive medical review and could lead to more safety/medical queries being generated.

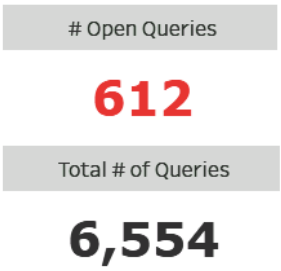

Figure 3

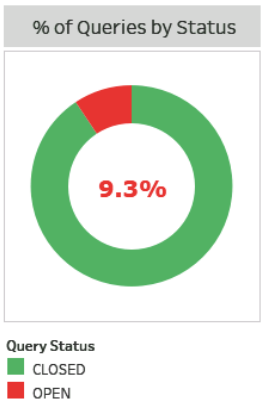

Figure 4

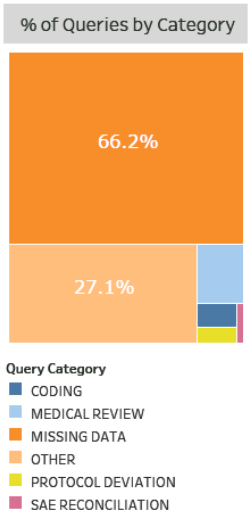

Figure 5

DETAILED QUERY STATUS

The query dashboard should be dynamic in order to answer more specific questions than only having a view on the overall status for a study. The more flexible it is, the better. In that way, all stakeholders could find the information they are looking for. It will help them monitoring a subset of the study population (e.g. a region, a country, a site, or even a specific subject).

EVOLUTION OVER TIME

The study team could be willing to see the current status of queries or to have a view on its evolution over time (see Figure 6). The dashboard can be adapted by selecting a specific month of interest. In most of the cases a decrease in queries resolution activities will be seen during holiday season. This information will be considered when setting-up study milestones calendar by allowing more time to sites for reaching their goals during those periods.

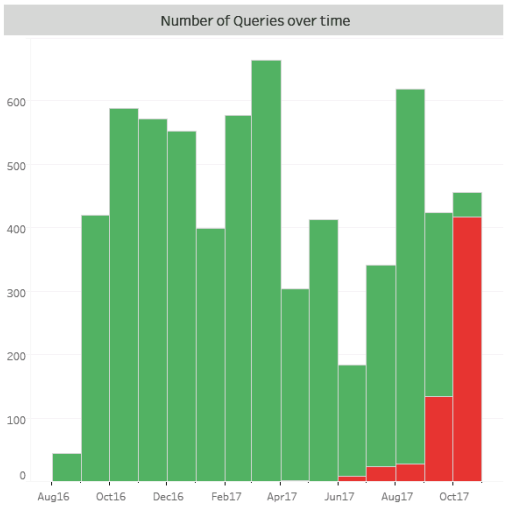

Figure 6

GEOGRAPHICAL DISTRIBUTION

The visualization should also give information on how the different countries and sites are performing (see Figure 7). Having visuals enabling the comparison between countries or between sites helps prioritizing activities. For example, the study could be managed by different regional clinical CROs. If it is noticed that query resolution in a particular region is below the expectations, actions could be defined to improve the situation (e.g. additional regional or country milestones to closely follow-up the status by defining intermediate deadlines, additional monitoring visits, country/site teleconference to go through the most problematic issues and provide further guidance,…)

Sometimes the study recruitment is not equally distributed among the participating countries. As a result, if your top recruiting countries are below the average time for queries resolution, this could seriously impact the global status of the study. By focusing on having those countries resolving most of their queries, one immediately sees a significant benefit on the study. All studies can't be managed the same way, there is a need to adapt the way they are handled, considering a lot of parameters including the recruitment rate. Usually the clinical operations team wants to see the number of queries by countries but also the average number of queries by patient so that they are not biased by the recruitment rate and can compare information. This could be provided at country or at site level (see Figure 8).

Figure 7

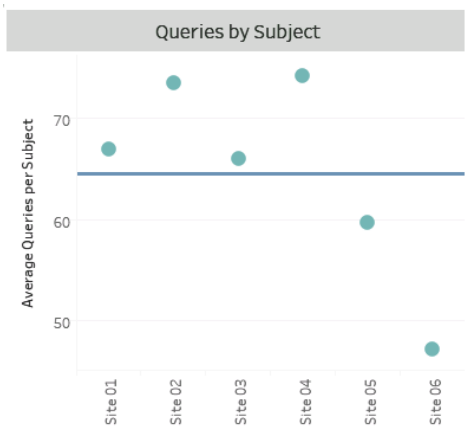

Figure 8

SITE STATUS

In large studies, the data cleaning could be the responsibility of several data managers and medical monitors. All of them will be interested by the sites they are in charge of. During the course of the study and especially at study milestones like interim/final analysis, they are chasing the sites to resolve all their queries within the defined timelines. Having the option to drill down at site level is therefore very important in their activities. Clinical operation manager and study project manager are, interested like for the countries, in the sites with the majority of the unanswered queries (see Figure 9). If those are targeted, an overall decrease in the number of outstanding queries will be achieved.

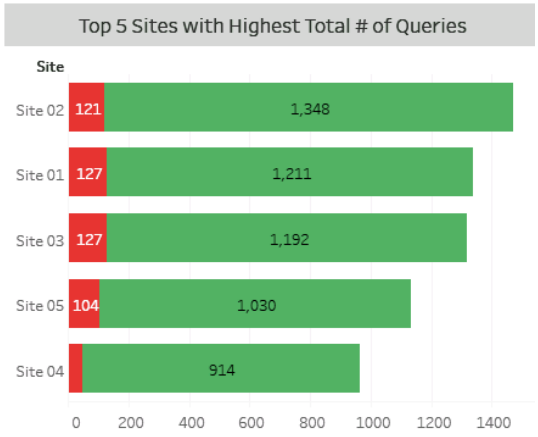

Figure 9

IDENTIFICATION OF TRENDS

SITES THAT ARE LESS RESPONSIVE

By looking at the average number of open queries by patient (see Figure 8) you can easily identify the centers that are less responsive. By filtering on one of those sites, you can see if the queries are from a specific category or if they are rather distributed equally within all categories. If a site is for example not answering the SAE and/or medical queries, it could mean that they need further guidance or even re-training on protocol requirements or eCRF completion.

Another interesting indicator could be the number of days since a query has been issued (see Figure 10). If the number of queries unanswered for more than one month is high, this might be due to lack of resources at the site. This can be further discussed with the Sponsor team. If the number of queries outstanding for a long time is very limited it could mean that the site needs support in replying queries on specific items. Maybe the site's staff thinks that issues have already been addressed and is therefore ignoring those questions or they simply don't understand what is requested by data manager/CRA/medical monitor. As soon as a query stays in the listing for several months, it should be a warning for the person who created it. Indeed he/she should first check that the query is well/still applicable. Then he/she should ensure that it is formulated in a clear way by indicating what the problem with current data is and by providing clear instructions on the different options for answering.

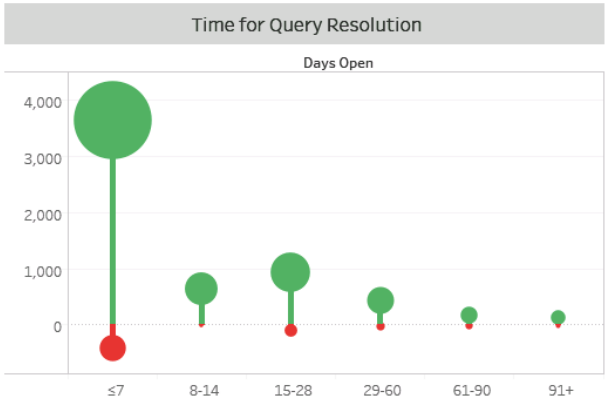

Figure 10

TYPE OF QUERIES RESULTING IN POOR RESOLUTION

In a study, not all queries will have the same level of importance. Clearly all questions related to patient's safety should always be prioritized. Indeed it is easily understandable that when a patient is experiencing an adverse event which is requiring study drug discontinuation as per protocol, the related questions are of utmost importance. Usually those queries are identifiable via query categories that alert site's staff that they must look at them in priority (see Figure 11). The dashboard should allow filtering on such sensible categories, like SAE, to have in few seconds the distribution of those queries within the different countries and sites.

It might also be interesting to know which visit or specific eCRF forms are generating most of the queries (see Figure 12). There will be no surprise that the screening visit (and the first subsequent visits) will be generating a huge amount of queries. Then we usually see a learning curve when the site staff is becoming familiar with the study protocol and also with the EDC system which could be new to them. Then the study treatment administration could also require additional cleaning efforts, depending on the complexity of the study design and the drug dispensing. And finally, the concomitant medications and adverse events are the main point of attention to understand the potential side effects of the study drug. As a result, those visits and forms are intensively explained during the investigators and monitors training on eCRF/EDC system. If any other visit or assessment is showing an unusual number of queries, it could be the starting point for a team discussion on whether protocol and/or CRF completion guidelines are clear enough and covering all situations. As an example, in Figure 12 we notice that there are a lot of queries at cycle 5. Upon investigation it was found that the majority of these queries were linked to the echocardiogram assessment. Indeed the protocol is requiring a left ventricular ejection fraction (LVEF) assessment to be performed with echocardiogram at cycle 5. However most of the sites are performing multiple-gated acquisition (MUGA) scan instead of echocardiogram. The eCRF being reflecting the protocol will request information from echocardiogram which will not be completed if another method was used. It will be generating several queries. The study team could decide that MUGA is an acceptable method and have the protocol/CRF completion guidelines amended accordingly.

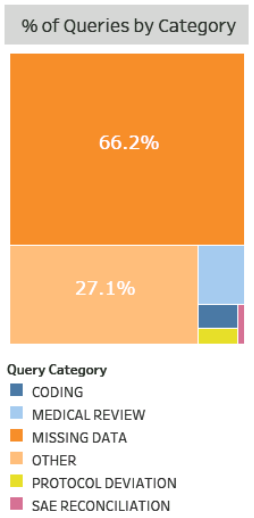

Figure 11

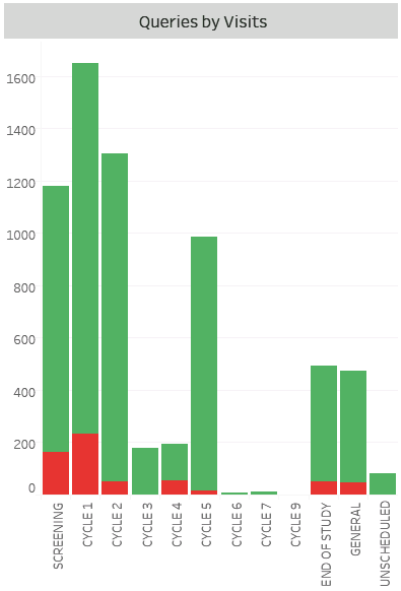

Figure 12

SITE OR QUERY LEVEL

To ensure that the dashboard is as complete as possible, it must give the possibility to go to the patient or even the query level. After having filtered on the different parameters that are of interest (e.g. a specific site, a category of query, query open for at least 30 days), the data manager or the monitor will be keen to see the queries details, including the query text (see Figure 13). For that purpose the relevant queries information should be shown as a table and be adapted to the visuals filters. In our example, the queries information table is shown at the bottom of the dashboard. It is a combination of the XQ dataset and the demographic information (DM dataset).

Queries Table

Study	Country	Site	Subject	Query Category	Visit	Query Text	Creation Date	
STUDY001	DEU	Site 06	STUDY001_06-007	PROTOCOL DEVIATION	SCREENING	Please complete Exclusion Criterion 11.	2016-10-12	CLOSED
						Please complete Exclusion Criterion 12.	2016-10-12	CLOSED
				SAE RECONCILIATION	GENERAL	According to the confirmed Outcome of this serious Adverse Event (not resolved), please consider to ..	2017-09-13	OPEN
	ITA	Site 02	STUDY001_02-001	CODING	GENERAL	To allow proper coding, could you please precise the active ingredients	2017-03-07	CLOSED
			STUDY001_02-002	CODING	GENERAL	To allow proper coding, could you please precise MH ID#, thank you.	2017-03-07	CLOSED
			STUDY001_02-005	SAE RECONCILIATION	GENERAL	There is a discrepancy in the term reported between Safety database ("Left lobe new pneumonia") a..	2017-03-14	OPEN
			STUDY001_02-006	CODING	SCREENING	To allow proper coding, could you please consider to split this term in two? Thank you.	2017-02-01	CLOSED
				PROTOCOL DEVIATION	SCREENING	Please complete Exclusion Criterion 1.	2017-04-10	CLOSED
						Please complete Exclusion Criterion 3.	2017-04-10	CLOSED
						Please complete Exclusion Criterion 4.	2017-04-10	CLOSED
			STUDY001_02-007	SAE RECONCILIATION	GENERAL	There is a discrepancy in the term between Safety database ("Pneumonia aggravated") and EDC ("...	2017-03-14	CLOSED
			STUDY001_02-010	SAE RECONCILIATION	GENERAL	There is a discrepancy in the severity reported between Safety database ("Moderate") and EDC ("Se..	2017-03-14	OPEN
			STUDY001_02-011	SAE RECONCILIATION	GENERAL	Start date of this serious Adverse event is after last dose of study drug (C2D21 27Mar2017), therefo..	2017-10-09	OPEN
						The SAE "EPILEPTIC SEIZURE" has been entered in eCRF but is missing in the Safety database. Pleas..	2017-06-30	OPEN

Figure 13

CONCRETE EXAMPLE OF USE AND INTERPRETATION

The queries dashboard can be filtered on the data of interest to quickly identify trends and take appropriate actions. As an example, we first select the "Open queries" from the chart on the left side of the dashboard (see Figure 14).

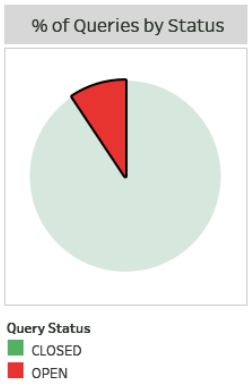

Figure 14

As a result, the dashboard will be showing the details of the open queries (see Figure 15).

Figure 15

One can see that most of the queries are outstanding in Romania (country colored in dark blue in 'Geographical Repartition'), that there are open queries that were already issued in March 2017 (in graph 'Number of Queries over time'), that there are still open queries for all visits (in graph 'Queries by Visit'), that site number 06 has the highest number of open queries by enrolled patient (in graph 'Queries by Subject'), …

You might then want to look into the queries that are outstanding for more than 3 months by selecting the category "91+" in the 'Time for Query Resolution' graph (see Figure 16).

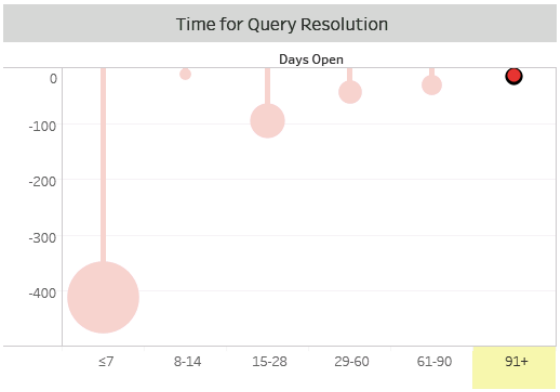

Figure 16

The dashboard will now be showing the details of queries outstanding for more than 3 months (see Figure 17).

Figure 17

In the treemap '% of Queries by Category', you can see that those queries are about missing data, coding issues and SAE reconciliation problems. This time, site number 01 is above the average number of queries by subject. When

you hoover on the graph, details are displayed in the tooltip and you can see that site 01 has an average of 1.6667 queries (see Figure 18) compared to the study average of 1.1333 queries (see Figure 19).

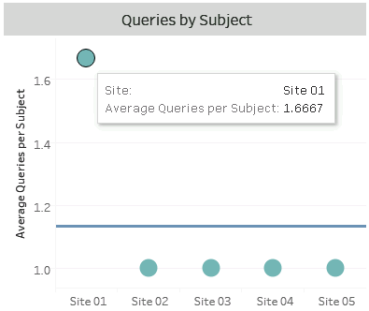

Figure 18

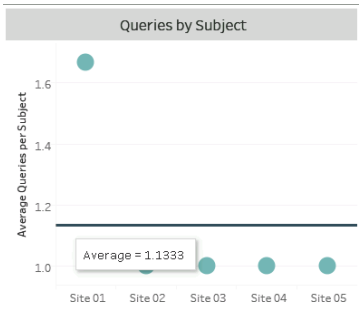

Figure 19

Those queries are displayed in the 'Queries Table' at the bottom of the dashboard.

You can filter the dashboard in many ways: by country, by site, by status, by category of queries, by time from creation… Parameters can also be combined.. For example several countries or several queries categories can be selected at the same time to retrieve the information you are looking for.

CONCLUSION

The use of a query dashboard could significantly improve study and site management. Developing this visualization with an SDTM-like model is allowing re-use across studies and various EDC system. The perspective from all end users should be considered to ensure that the dashboard will be as flexible as possible and will meet defined goals.

REFERENCES

1. INTEGRATED ADDENDUM TO ICH E6(R1): GUIDELINE FOR GOOD CLINICAL PRACTICE (R2) – Dated 9 November 2016
   - *http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6/E6_R2__Step_4_2016_1109.pdf*
2. FDA Guidance for Industry - Oversight of Clinical Investigations – A Risk Based Approach to Monitoring
   - *https://www.fda.gov/downloads/Drugs/Guidances/UCM269919.pdf*

ACKNOWLEDGMENTS

We wish to thank our colleagues who generously provided their insight and perspective, that greatly helped to improve the queries dashboard and the quality of this paper: Benjamin Devriendt, Jean-Noël Wallet, Roxane Debrus and Nick De Donder.

CONTACT INFORMATION

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

Joke De Wever
Business & Decision Life Sciences
Sint-Lambertusstraat 141
1200 Brussels
+32 485 731058
joke.dewever@businessdecision.be
https://www.businessdecision-lifesciences.com/

Sabine Jonas
Business & Decision Life Sciences
Sint-Lambertusstraat 141
1200 Brussels
+32 496 266410
sabine.jonas@businessdecision.be
https://www.businessdecision-lifesciences.com/

Irina Sargsyan
Business & Decision Life Sciences
Sint-Lambertusstraat 141
1200 Brussels
+32 491 868909
irina.sargsyan@businessdecision.be
https://www.businessdecision-lifesciences.com/