medidata | WHITE PAPER

Medidata Detect:

A Vital Tool for Monitoring eCOA/ePRO Data to Improve Data Quality and Reduce Trial Risk

Table of Contents

Introduction

Life sciences companies are simultaneously facing three separate but related forces that will shape the way that they collect and monitor data on the patient/caregiver experience in clinical trials.

First, in accordance with the 21st Century Cures Act, the US Food and Drug Administration (FDA) is developing a series of guidance documents to address the use of patient/caregiver data in regulatory decision making.[1] Thus, Sponsors are actively working to represent the voice of the patient in their regulatory submissions and increasingly are collecting electronic Clinical Outcomes Assessments (eCOA) and/or electronic Patient-Reported Outcomes (ePRO) in their clinical trials.

Second, ICH E6 (R2) which was adopted in December 2016 has encouraged sponsors to adopt a more proactive approach to study design, risk management and study monitoring, especially risk based monitoring. Many regulators have produced guidance documents, including the FDA which has most recently published a Question-and-Answer document intended to facilitate Sponsors' implementation of Risk-Based Monitoring (RBM). This Q&A document further supports the agency's belief that "risk-based monitoring is an important tool to allow sponsors to identify and address issues during the conduct of clinical investigations."[2]

Third, the International Council for Harmonization (ICH) has published draft revisions to ICH E8 (R1) to address challenges with designing and executing trials with diverse designs and data sources. The revision emphasizes that the quality of clinical research — required to protect patients and generate reliable and meaningful data — relies on both good *design* and good *execution*. This revision also stresses the importance of Quality by Design (QbD) approach to clinical trials, and the need to rapidly identify and address systemic issues.

As the industry moves to adopt patient-centric approaches to clinical trials, there is an opportunity to adopt the best practices of QbD and risk-based methodologies to ensure that patient reported data are effectively being monitored.

In this paper, we discuss the value of using advanced technology — specifically centralized statistical analytics — to oversee subject safety and data quality/integrity in a risk-based manner. Such monitoring is particularly relevant for the monitoring of large data sets, such as eCOA/ePRO data, as all regulators agree that a risk based approach including Centralized Statistical Monitoring of data in real time may enable sponsors to be able to identify the reasons for missing data, inconsistent data, protocol violations, or protocol deviations etc. and take corrective actions to minimize the likelihood of recurrence for the remainder of the clinical investigation.

Despite the scientific rigor that the pharmaceutical industry has applied to generating quality data throughout the clinical trial process, data quality issues still arise — sometimes with drastic consequences. Nearly a quarter (24 percent) of applications to regulators require one or more resubmissions before approval. Over half (52 percent) of resubmissions that are eventually approved resulted from inconsistent study results, either across endpoints, sites, or studies. The first unsuccessful submission adds a median delay of 435 days to the approval process.[3] With considerable financial cost, these unsuccessful submissions and delays represent significant waste and in incalculable loss for all stakeholders.

RBM: A Component of Risk-Based Quality Management

Risk-based monitoring (RBM) serves as a risk control mechanism to oversee data quality during study conduct, including centralized review. The goal of RBM is to use resources efficiently in ensuring patient safety and data quality, minimizing the need for costly and time-consuming visits to investigator sites to perform verification of 100 percent of trial data. The FDA's guidance on the subject, "A Risk-Based Approach to Monitoring," was published in 2013 and stresses that "risk-based approaches to monitoring, focused on risks to the most critical data elements and processes necessary to achieve study objectives, are more likely than routine visits to all clinical sites, and 100 percent data verification to ensure subject protection and overall study quality."

RBM is one component of Risk-Based Quality Management (RBQM), which is a process for planning and executing clinical trials that is rooted in the principles of QbD. The subject of regulatory reflection papers and countless industry articles, RBQM is beyond the scope of this paper, other than to note that it is an overarching framework for preventing, monitoring, and mitigating risk in trial planning and execution. It begins with protocol design and runs all the way through study closeout and should support key decisions and manage clinical trial complexity.

Performing RBQM with Centralized Statistical Analytics of eCOA/ePRO Data

As mentioned earlier, it is virtually impossible to sufficiently monitor eCOA/ePRO data unaided by computer analytics due to the volume involved. Specifically, the job calls for a tool that uses machine-learning algorithms and high-dimensional analytics to highlight trends and relationships and to reveal anomalies within the data. To function fully, the tool should be part of an integrated platform that cleans and manages data and is used across all monitoring functions to identify, document, and manage Key Risk Indicators (KRIs). Here, however, we are focused only on its application to eCOA/ePRO data.

How does it work? Medidata Detect is a centralized, statistical analytics software that "learns" the proper, acceptable ranges for all data fields. It then combs millions of data points, comparing every variable in the data set to every other variable, looking for and identifying statistical relationships between those variables. It identifies thousands of patterns in the data and spots inconsistencies or outliers that don't fit the pattern or that fall outside of the acceptable ranges established. It then flags values related to patients or sites that should be investigated and possibly remedied. Medidata Detect focuses on areas that are expected to be problematic, i.e. words those known risks defined before the study's start, but is able to detect unusual patterns or values that arise even if they relate to unknown and unpredictable risks.

A best-in-class centralized statistical analytics system:

- Refreshes the data in real time
- Is data source agnostic, allowing the Sponsor to define the data sources to be monitored
- Allows the Sponsor to define the Key Risk Indicators (KRIs)
- Requires little set up so the Sponsor does not have to configure the algorithms Provides
- data listings, summary tables, patient profiles, scatter plots and box plots

Detectable Issues in eCOA/ePRO Data

NONCOMPLIANCE WITH THE TREATMENT REGIMEN

Typically, it is difficult to track a patient's compliance with the treatment regimen that is specified in the protocol using standard data collection. Doing so becomes easier with the use of electronic tools — especially when centralized statistical analytics are applied to it. It is possible to detect on a daily basis if patients are taking their medication as directed.

TIME/DATE ERRORS

Measures provided by patients via electronic tools are never "wrong" per se, in that they reflect the patient's experience. However, patient data may be reported incorrectly or incompletely. For example, the date and time of recorded data might not make sense as in the example shows below (see Figure 1). In this example, two patients have recorded responses to a quality of life questionnaire (EQ5D) after their date of death. Obviously this warrants further investigation.

Figure 1: Example of Illogical Dates of Data Entry

FS1369									
Date	**Visit**	**Anxiety/Depression**	**Mobility**	**Pain/Discomfort**	**Self-Care**	**Usual Activities**	**VAS Score**	**Ques. Admin**	**Completion Status**
13-Dec-2017 12:12	Day 1 Randomization	I am not anxious or depressed	I have severe problems walking	I have moderate pain or discomfort	I have slight problems washing or dressing myself	I have moderate problems doing my ususual activities	70	Self-administered	Completed
04-Apr-2018 12:33	Visit 2							Interviewer administered	Not completed due to site staff error
08-Aug-2018 14:06	Visit 3	I am not anxious or depressed	I have slight problems walking	I have no pain or discomfort	I have no problems washing or dressing myself	I have slight problems doing my ususual activities	91	Self-administered	Completed
12-Oct-2018 12:45	Visit 4	I am not anxious or depressed	I have slight problems walking	I have slight pain or discomfort	I have moderate problems washing or dressing myself	I unable to do my ususual activities	60	Self-administered	Completed
02-Feb-2019 11:32	Visit 5							Interviewer administered	Other
06-May-2019 14:23	Visit 6							Interviewer administered	Subject died
15-Aug-2019 10:05	Visit 6							Interviewer administered	Subject died
21-Oct-2019 13:11	Visit 6							Interviewer administered	Subject died

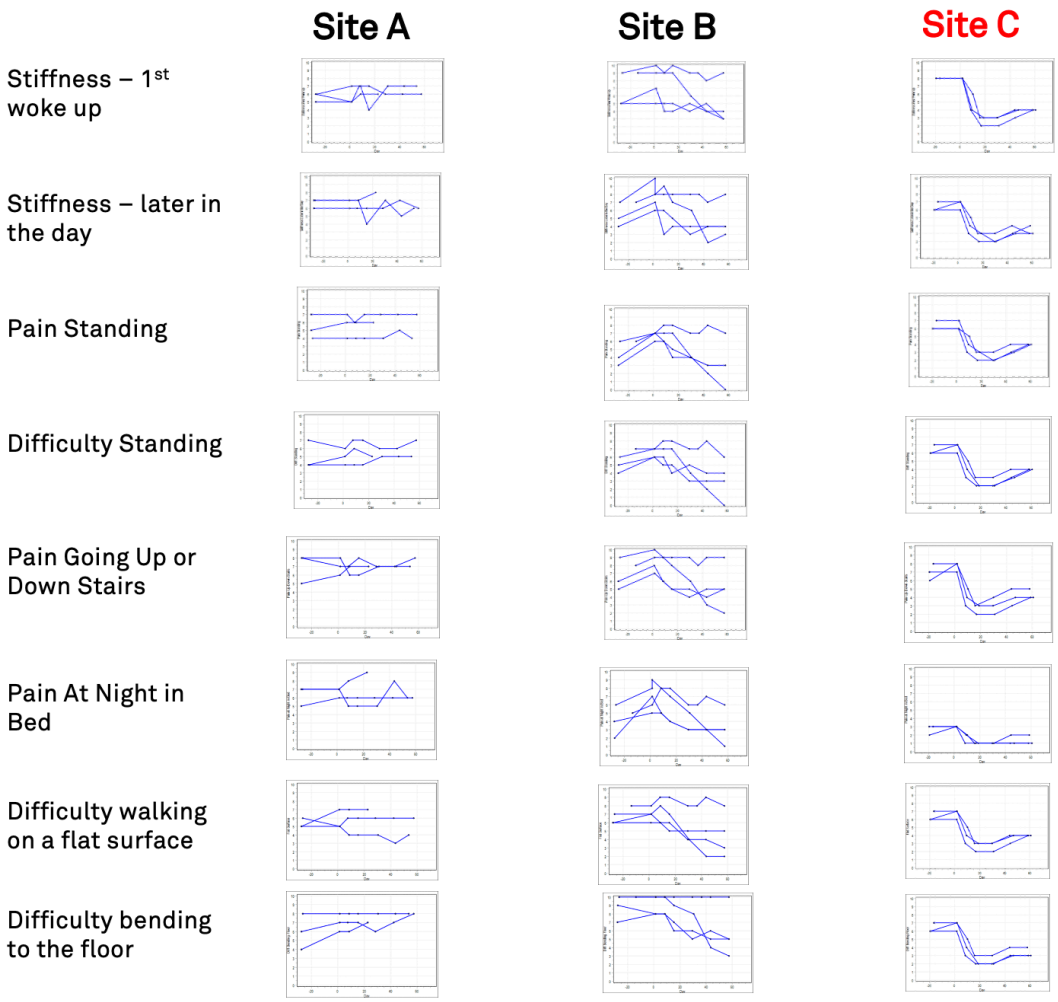
MISSING DATA

Patients can, of course, neglect to make a diary entry — a situation that can be identified by Medidata Detect and that Sponsors will want to address if it becomes a pattern.

ISSUES WITH REPORTING INSTRUCTIONS

Sites can misinterpret how patients should use their diaries and inadvertently mislead patients when giving them instructions. When this is the case, answers from patients at such a site will be different from those at others. In Figure 2, patient responses from Sites A and B show normal variations. However, those from Site C are suspiciously uniform, which should raise a concern about how patients are reporting their information.

Figure 2: Example of Questionable Site Data



DATA FABRICATION

Although it is extremely rare, there are instances of misconduct in which a site might complete patients' own assessment. Centralized statistical analytics are able to catch this by identifying patterns in the data that seem "off."

INCONSISTENCIES BETWEEN CLINICIAN AND PATIENT REPORTS

It is possible to detect when a physician's assessment of a patient's progress is markedly different from the patient's. In the example shown in Figure 3, the clinician and subject assessments have wildly different trajectories over more than 700 days of reporting. The physician's assessment on the right indicated a precipitous decline and then a sustained low level of disease activity, whereas the patient's assessment of disease and pain remained high, and even increased over the same period. This could suggest that the scales used in the patient's responses have been consistently misinterpreted.

Figure 3: Example of Incongruous Patient and Physician Assessments

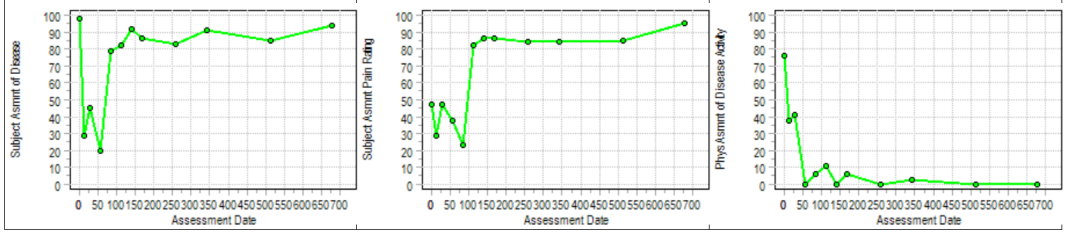
Patient 10091001

PROFILE | ISSUES | QUERIES | DOWNLOAD PDF | DOWNLOAD WORD

VAS Efficacy Outcome

Date	*Visit Name*	*Assessment Administered*	*Subject Asmnt of Disease*	*Subject Asmnt Pain Rating*	*Phys Asmnt of Disease Activity*	*Phys Asmnt Change from Baseline*	*ALT (U/L)*
08Feb2011	Baseline	Yes	98	47	76	0	15
22Feb2011	Week 2	Yes	29	29	38	-38	26
08Mar2011	Week 4	Yes	45	47	41	-35	24
05Apr2011	Week 8	Yes	20	38	0	-76	27
03May2011	Week 12	Yes	79	23	6	-70	24
31May2011	Week 16	Yes	82	82	11	-65	21
28Jun2011	Week 20	Yes	92	86	0	-76	21
26Jul2011	Week 24	Yes	86	86	6	-70	18
25Oct2011	Week 37	Yes	83	84	0	-76	
17Jan2012	Week 49	Yes	91	84	3	-73	23
04Jul2012	Week 73	Yes	85	85	0	-76	17
18Dec2012	Week 97	Yes	94	95	0	-76	23

VAS Efficacy Graphs



REAL-TIME ANALYSIS, TIMELY INTERVENTION, AND REDUCED RISK

Sponsors interested in capturing the patient voice in trials via eCOA/ePRO can now do so with confidence in their ability to oversee the incoming data properly through centralized statistical analysis. Medidata Detect identifies data anomalies and process deficiencies that would likely go undetected without its innovative approach using sophisticated computerization.

Because the data are available in real time, the tool supports timely issue identification and resolution. Quality checks are performed automatically as the data come in, and potential problems are flagged at once. *This gives monitors the opportunity to investigate and take corrective action before a problem is replicated.* Electronic patient sourced data, once submitted, cannot be changed or corrected. The only recourse is to prevent errors from becoming systemic.

Implementing Medidata Direct makes it far less likely that data from a given patient or site would be deemed invalid and not submissible to regulators. Indeed, the FDA encourages such early intervention:

> "There may be situations in which poor trial conduct or adherence to the investigational plan causes or contributes to incomplete data collection. Therefore, by reviewing important investigation activities, in real-time across CI sites, sponsors may be able to identify the reasons for missing data, protocol violations, or protocol deviations and take corrective actions to minimize the likelihood of these occurring during the remainder of the clinical investigation."[3]

The systematic efficiency with which the software performs the monitoring function reduces both the number of pre-programmed edit checks and custom reports that are necessary, as well as the associated costs. Consequently, it will help those responsible for data quality fully embrace regulations while focusing their work on higher-value activities — those significant items that require immediate attention.

And by identifying anomalous data early, Medidata Detect gives Sponsors the enormous advantage of being able to resolve issues as they arise, which reduces the cycle time between the last patient's last visit (LPLV) and database lock, ultimately speeding the time to filing a new drug application (NDA).

Overall, monitoring for eCOA/ePRO data quality using Medidata Detect increases the probability of the study's success by reducing the risk that subject data will need to be excluded or that the study will require additional regulatory reviews. Figure 4 highlights how centralized statistical analytics transforms data management by reducing risk, cycle time and cost.

Figure 4: Value of Centralized Statistical Analytics

CSA Transforms Data Management Reducing Risk, Cycle Time and Cost

Reduce Edit Checks	Automate Data Review	Reduce Custom Reports	Faster Database Lock	Reduced Study Risk
Pattern recognition and outlier detection to reduce number of pre-programmed edit checks	Automated data review by flagging data anomalies using machine learning and advanced statistical analysis	Reduce number of custom reports created and run	Early identification and resolution of anomalous data	Reduced risk and increased probability of successful study
Reduce associated cost Requirements gathering, logic development, programming, validation, project management	Data managers are able to focus efforts on significant items that require attention	CSA provides data listings, summary tables, patient profiles, scatter plots & box plots out of the box	Reduction in LPLV to DBL cycle time resulting in earlier filing of NDA	Increased confidence in study findings
			Reduced number of database unlocks after initial database lock	Reduce risk subject data excluded from study
				Reduce risk of additional regulatory review

Conclusion

Medidata Detect gives Sponsors the ability to operationalize a patient-centric approach, folding patient input into their drug development plans. Relying on centralized statistical analytics will help ensure the quality of eCOA/ePRO data and compliance with ICH and FDA regulations on the use of risk-based monitoring. Certainly, studies collecting eCOA/ePRO should be using Medidata Detect, as it is simply not humanly possible to spot patterns in the data without the aid of advanced analytics. The power of Medidata Detect is in allowing CRAs the ability to perform root-cause investigations and take corrective actions, rather than doing the work that machine learning computers can perform.

By monitoring eCOA/ePRO data in real time, Medidata Detect identifies quality issues at once so that they can be addressed early, before they are repeated or become systemic and jeopardize the integrity of the study.

Medidata Detect is a foundational step along the path to adopting a RBQM approach. The capabilities of Medidata Detect will grow even more critical as trials shift toward greater degrees of virtualization. As protocols increasingly depend on data collection of patient symptoms and response outside of the clinic, it will be ever more important to spot potential problems quickly — in time to intervene.

Endnotes

1. https://www.fda.gov/drugs/development-approval-process-drugs/fda-patient-focused-drug-development-guidance-series-enhancing-incorporation-patients-voice-medical

2. "A Risk-Based Approach to Monitoring of Clinical Investigations Questions and Answers Guidance for Industry," FDA, March 2019

3. Sacks LV, Shamsuddin HH, Yasinskaya YI, Bouri K, Lanthier ML, Sherman RE, "Scientific and Regulatory Reasons for Delay and Denial of FDA Approval of Initial Applications for New Drugs, 2000-2012." *JAMA*.2014;311(4):378–384. doi:10.1001/jama.2013.282542

About Medidata

Medidata is leading the digital transformation of life sciences, creating hope for millions of patients Medidata helps generate the evidence and insights to help pharmaceutical, biotech, medical device and diagnostics companies, and academic researchers accelerate value, minimize risk, and optimize outcomes. More than one million registered users across 1,400 customers and partners access the world's most-used platform for clinical development, commercial, and real-world data.

Medidata, a Dassault Systèmes company (Euronext Paris: #13065, DSY.PA), is headquartered in New York City and has offices around the world to meet the needs of its customers. Discover more at **www.medidata.com** and follow us **@medidata**, The Operating System for Life Sciences™.

info@medidata.com | +1 866 515 6044