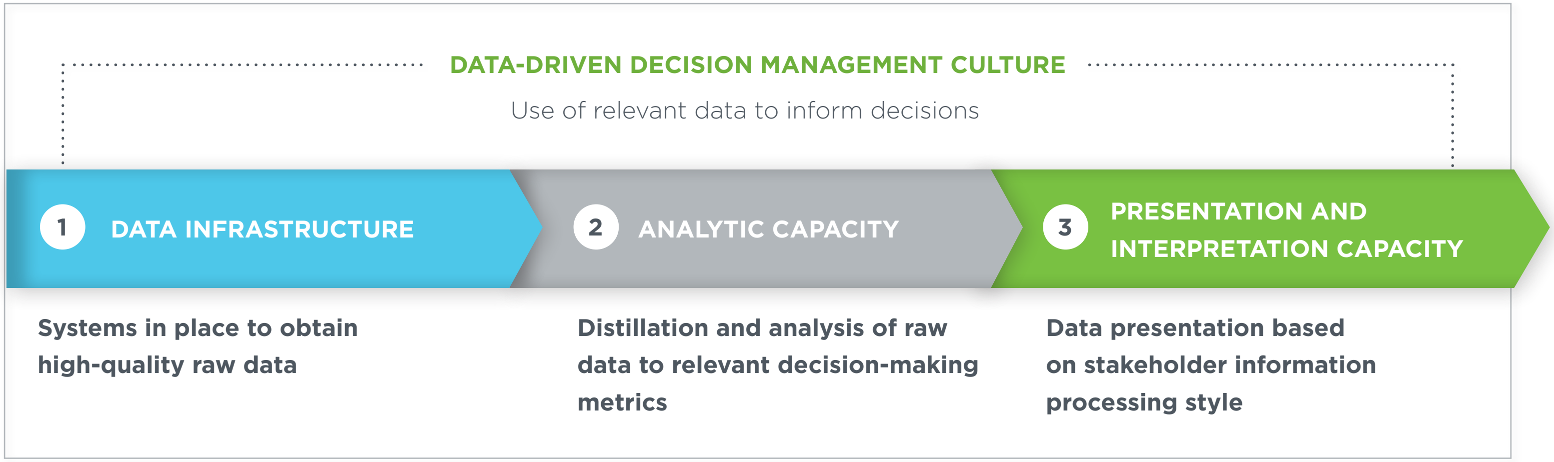
Data-Driven Decision Management (DDDM) in Clinical Trials: THE WAVE OF THE FUTURE IS HERE

A Bioclinica White Paper

Data-driven decision management (DDDM) is an approach to business governance that values decisions supported by verifiable data. The DDDM process transforms data into useable knowledge that can be used to improve business outcomes. It is equally effective for decisions both big and small and can be applied across a wide range of industries. In fact, many organizations are already embracing the need to leverage data in their decision-making processes. It can be seen in practice when a major league baseball team uses statistical evidence to alter a pitcher's throwing motion to prevent injury. It is also exemplified when key performance indicators (KPIs) are used to re-allocate marketing dollars for patient recruitment in a clinical trial because one marketing channel was identified as more effective than others. Study results from the MIT Center for Digital Business indicate that when DDDM is utilized, organizations benefit from 4% higher productivity rates and 6% higher profits compared to their competitors.

DDDM is the wave of the future, and the future has arrived. If metrics, statistics and KPIs are not a big part of your organization's vocabulary, then you are probably not riding on the DDDM wave. But not to worry – it is not too late to move your organization into the future. The simple framework in Figure 1 is all it takes to get started.

FIGURE 1: Data-Driven Decision Management (DDDM) Framework



In each of the following sections, we will describe the requirements for each step in the DDDM process and some of the common challenges to overcome to achieve success.

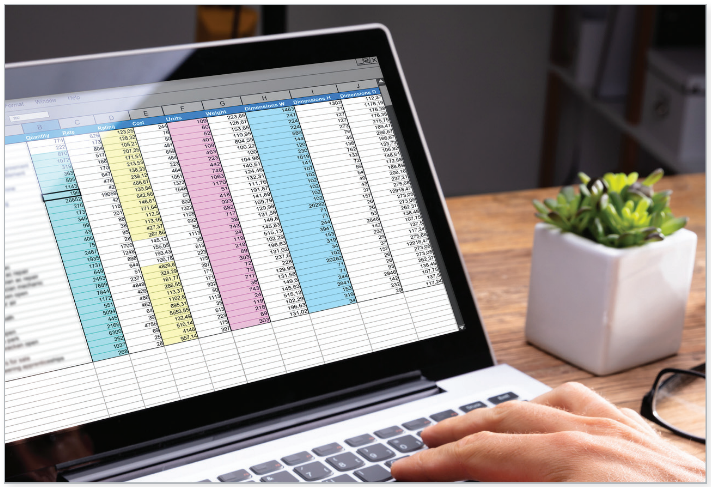
Obtaining High-Quality Raw Data

To implement DDDM, the first step is having the data infrastructure in place to obtain high-quality raw data because having the right data is paramount to driving effective decisions. To be high quality, the data must be accurate, complete, timely and actionable. The data should also be well defined, easily identified, accessible and compatible across data systems so multiple users can readily view and utilize it.

Depending on the maturity of the data infrastructure systems within your organization, there could be several unique challenges to obtaining high-quality raw data.

Organizations with less mature data infrastructures can face challenges with both data completeness and quality. If technology is not available to automate data acquisition for decision-making purposes, manual data collection processes might be required. As an example, a technology organization wanting to improve system performance may decide to use DDDM to guide where to apply limited investment dollars. To determine where to focus development efforts for the greatest impact on user experience, a key input is the response times for all application functions. If automated scripts are not available to test response times, testers will have to use a stopwatch approach, during which the response times of each application function are manually recorded. This is both time consuming and susceptible to error, which can decrease data quality. Although manual data acquisition can be feasible, automation is most often the preferred and less error-prone approach. In this example, the risk of poor data quality from the stopwatch approach is too high, potentially driving nonoptimal investment decisions. Instead, it would be worthwhile for the organization to invest in automated testing scripts to strengthen their data infrastructure.

On the opposite end of the spectrum, organizations with mature data infrastructures could be challenged with data overload. The data they need to drive decisions is available, high quality and obtained in an automated fashion. But, as their databases grow, finding the right data can be like finding a needle in a haystack. These organizations will often have to employ a team of Data Scientists experienced in Big Data to isolate the right data and drive their decisions. Implementing a data infrastructure with this level of sophistication requires significant investments, but the return on investment (ROI) from optimal decision making will often outweigh the costs. The success of organizations like Google, whose entire business model depends on using vast amounts of data for business decisions, demonstrates that mature data infrastructures can pay big dividends.



CASE STUDY

Obtaining High-Quality Raw Data for Clean Patient Tracking

Situation

A key step in locking a clinical trial database is ensuring that clean data is obtained for all patients participating in the study. To do this effectively and with confidence, high-quality raw data is needed to make optimal data-driven decisions (DDDM) regarding clean patient data status.

The following needs to be achieved before considering a patient as clean:

- No open queries for the site to respond to or for the Clinical Data Manager (DM) or Clinical Research Associate (CRA) to review and close
- No missing pages or visits
- Completed source data verification, as appropriate
- All terms medically coded, as appropriate
- All programmed DM data review listings completed
- All data reconciled that are external to the Clinical Data Management System (CDMS), such as serious adverse events (SAEs), safety labs, ECGs, imaging and ePRO

The amount of data that you must review and reconcile to achieve clean patient data status can be daunting, particularly when multiplied across all study sites and patients. This is where a programmatic approach to tracking clean patient data can be useful and accelerate database lock.

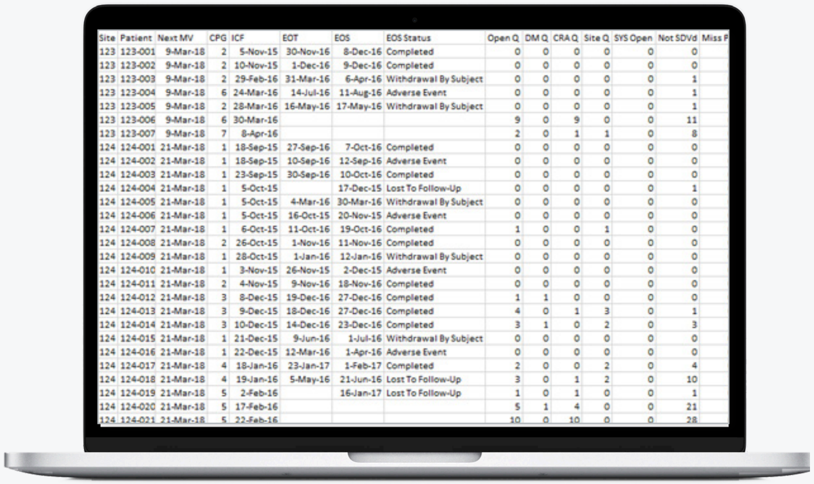
Solution

To assist with obtaining high-quality raw data, a clinical data management tool that aggregates eSource data from multiple sources (EDC, ECG, safety labs, ePRO) into one comprehensive source database could be useful. Ideally, this system would have built-in validation, further reducing the workload. This type of tool would facilitate optimal decisions about clean patient data status.

Outcomes

The data aggregated within the tool would provide Clinical Data Management teams with the data they need to target critical path open tasks for cleaning patient data, track trends in task completion and provide executive-level status to management on progress towards database lock.

FIGURE 2: Example of an aggregated data source to track clean patients



Site	Patient	Next MV	CPG	ICF	EOT	EOS	EOS Status	Open Q	DM Q	CRA Q	Site Q	SYS Open	Not SDVd	Miss F
123	123-001	9-Mar-18	2	5-Nov-15	30-Nov-16	8-Dec-16	Completed	0	0	0	0	0	0	
123	123-002	9-Mar-18	2	10-Nov-15	1-Dec-16	9-Dec-16	Completed	0	0	0	0	0	0	
123	123-003	9-Mar-18	2	29-Feb-16	31-Mar-16	6-Apr-16	Withdrawal By Subject	0	0	0	0	0	1	
123	123-004	9-Mar-18	6	24-Mar-16	14-Jul-16	11-Aug-16	Adverse Event	0	0	0	0	0	1	
123	123-005	9-Mar-18	2	28-Mar-16	16-May-16	17-May-16	Withdrawal By Subject	0	0	0	0	0	1	
123	123-006	9-Mar-18	6	30-Mar-16				9	0	9	0	0	11	
123	123-007	9-Mar-18	7	8-Apr-16				2	0	1	1	0	8	
124	124-001	21-Mar-18	1	18-Sep-15	27-Sep-16	7-Oct-16	Completed	0	0	0	0	0	0	
124	124-002	21-Mar-18	1	18-Sep-15	10-Sep-16	12-Sep-16	Adverse Event	0	0	0	0	0	0	
124	124-003	21-Mar-18	1	23-Sep-15	30-Sep-16	10-Oct-16	Completed	0	0	0	0	0	0	
124	124-004	21-Mar-18	1	5-Oct-15		17-Dec-15	Lost To Follow-Up	0	0	0	0	0	1	
124	124-005	21-Mar-18	1	5-Oct-15	4-Mar-16	30-Mar-16	Withdrawal By Subject	0	0	0	0	0	0	
124	124-006	21-Mar-18	1	5-Oct-15	16-Oct-15	20-Nov-15	Adverse Event	0	0	0	0	0	0	
124	124-007	21-Mar-18	1	6-Oct-15	11-Oct-16	19-Oct-16	Completed	1	0	0	1	0	0	
124	124-008	21-Mar-18	2	26-Oct-15	1-Nov-16	11-Nov-16	Completed	0	0	0	0	0	0	
124	124-009	21-Mar-18	1	28-Oct-15	1-Jan-16	12-Jan-16	Withdrawal By Subject	0	0	0	0	0	0	
124	124-010	21-Mar-18	1	3-Nov-15	26-Nov-15	2-Dec-15	Adverse Event	0	0	0	0	0	0	
124	124-011	21-Mar-18	2	4-Nov-15	9-Nov-16	18-Nov-16	Completed	0	0	0	0	0	0	
124	124-012	21-Mar-18	3	8-Dec-15	19-Dec-16	27-Dec-16	Completed	1	1	0	0	0	0	
124	124-013	21-Mar-18	3	9-Dec-15	18-Dec-16	27-Dec-16	Completed	4	0	1	3	0	1	
124	124-014	21-Mar-18	3	10-Dec-15	14-Dec-16	23-Dec-16	Completed	3	1	0	2	0	3	
124	124-015	21-Mar-18	1	21-Dec-15	9-Jun-16	1-Jul-16	Withdrawal By Subject	0	0	0	0	0	0	
124	124-016	21-Mar-18	1	22-Dec-15	12-Mar-16	1-Apr-16	Adverse Event	0	0	0	0	0	0	
124	124-017	21-Mar-18	4	18-Jan-16	23-Jan-17	1-Feb-17	Completed	2	0	0	2	0	4	
124	124-018	21-Mar-18	4	19-Jan-16	5-May-16	21-Jun-16	Lost To Follow-Up	3	0	1	2	0	10	
124	124-019	21-Mar-18	5	2-Feb-16		16-Jan-17	Lost To Follow-Up	1	0	1	0	0	1	
124	124-020	21-Mar-18	5	17-Feb-16				5	1	4	0	0	21	
124	124-021	21-Mar-18	5	22-Feb-16				10	0	10	0	0	28	

Data Distillation for Effective Analysis

Once an organization obtains a complete, high-quality set of data to aid their decision-making processes, the data may have to be further distilled, depending on its complexity, to be effectively analyzed. The data distillation process can take several forms, but most commonly it involves applying structure to unstructured data or removing data elements that are irrelevant to the decision-making process.

The previous examples of major league baseball and marketing for clinical trial recruitment, which were used to demonstrate the relevance of DDDM for all organizations, also lend themselves well to explaining the most common types of data distillation.

A decision to alter a major league pitcher's throwing motion to prevent injury would likely be based on data obtained from PITCH F/x. PITCH F/x is a camera-based system that is installed in every major league stadium and tracks the velocity, movement, release point, spin and pitch location for every pitch thrown. Automated systems obtain this data in real-time, but the data must subsequently be structured correctly to support decision-making processes. For decisions regarding changes to a pitcher's throwing motion based on PITCH F/x data, first, the data would have to be structured to identify other pitchers with similar release points. Second, the list of players with similar release points would have to be reconciled against injury history data for those same players. If injuries were more common for pitchers with this release point, then the data could support a decision to alter a pitcher's throwing motion to change their release point and hopefully reduce the likelihood of injury.

The decision to optimize a marketing budget for clinical trial patient recruitment can also be a byproduct of data distillation. An organization specializing in patient recruitment can have marketing effectiveness data spanning many clinical trials. A holistic look at this data may indicate that a specific marketing channel, on a particular day and time, would return the most effective recruitment results. However, if the data is distilled down to only clinical trials with an indication that aligns to the decision at hand, the decision process will yield a different and likely more accurate result.

CASE STUDY

Distilling Data for Effective Analysis of Clean Patient Status

Situation

As a study moves closer to database lock, to target specific tasks for cleaning patient data, the sponsor may want to identify the patients that have only one element preventing them from reaching clean status, what that one element is and which of their multiple vendors owns the element in question. Additionally, the sponsor management team may want to understand high-level metrics, such as the overall percentage of patients who have reached clean status, or to identify trends in the velocity with which certain key tasks are being completed. This data can help drive decisions around whether additional data management resources are needed to meet the target database lock date.

Depending on the size of the study and the frequency at which the data set is generated, the study's data set for clean patient tracking can grow quite large and become difficult to manage. While all the data elements in the data set for clean patient tracking are necessary, data distillation will often be required to answer questions such as those posed in the previous paragraph.

Solution

It would not unreasonable for a single aggregated data tool, as described in the previous case study, to contain thousands of rows and dozens of columns. To facilitate informed decisions based on this data, built-in filtering, formulas and functions could target and structure the relevant data.

Outcomes

Through a combination of target filtering of the aggregated data and the built-in formulas and functions, sponsors can easily distill the clean patient data down to manageable subsets for critical decisions related to database lock. For instance, the aggregated data can be filtered to identify patients that could become problematic for a timely database lock because their number of open tasks is greater than desired.

FIGURE 3: Target filtering in an aggregated data tool provides information about the number of tasks remaining.

Presenting Data for Optimal Interpretation

Once you have obtained high-quality raw data and it has been distilled for effective analysis, the final element of the DDDM process is to identify the ideal mechanism to present the data for optimal interpretation. The type of data being presented and the audience consuming the data are both critical components for selecting the best mechanism for presentation. Certain data may lend well to being presented in a report format, and other data may be better presented using graphs and other visuals. People are also unique in how they best process information. Some are visual learners, who best process and remember what they see. Others are verbal learners, who more easily process words, either written or spoken.

In the example of deciding whether to alter a major league pitcher's throwing motion to reduce their risk of injury, a visual approach to presenting the data could be the best option. The supporting data can be transformed into graphs that demonstrate the correlation of different release points and time spent on the disabled list. Additional visuals such as videos or still frames from the time of the injury that landed the pitcher on the disabled list could also be effective in convincing the pitcher to decide to change their throwing motion. Another argument for using the visual presentation approach is that people involved with athletics at a professional level are often skilled visual learners. They spend many hours dissecting game films to improve their performance and better understand their opponent's strengths and weaknesses.

In deciding how to optimize a marketing budget for clinical trial patient recruitment, a verbal approach to presenting the data could be the optimal choice. The data could be prepared in a detailed report format that guides stakeholders through the analytical process. The report might start with an analysis of marketing effectiveness data from all historical trials and then drill down to data for the indication and demographics for the trial in question. Business executives are often detail-oriented, analytical thinkers and can be uncomfortable with ambiguity. The detailed report format can arm them with the complete suite of information they require to be confident they are making an informed decision.

CASE STUDY

Presenting Data for Optimal Interpretation of Clean Patient Tracking Status

Situation

Based on my experience working with sponsors and their aggregated data, the data is most commonly being utilized in three ways:

1. To determine which tasks were outstanding to bring patients to a clean data status and who owned those tasks
2. To track trends in task areas to identify and remove any potential roadblocks to completion
3. To present high-level metrics to management that included measures such as total percentage of clean patients

Sponsors also want to generate visual dashboards based on the aggregate data to help understand the status.

Solution

For this purpose, a web-based visualization tool based on the aggregated data would be useful. The visualizations could update with the creation of each new aggregated data set, and the interface could contain dashboards for each of the three focus areas identified above. Operations teams can focus on the dashboards for clean data status and trends to support decisions to expedite database lock, while management teams can access the executive-level dashboard to stay up to date with overall progress and support decisions around resource optimization.

Outcomes

The visuals allow the operational and management teams to quickly ascertain the information they need to support their decisions, which is key because speed is paramount as sponsors move closer to database lock.

FIGURE 4: Example of a visual dashboard showing clean patient status based on the aggregated data

Creating a DDDM Culture

As we've described throughout this paper, implementing the DDDM framework is a straightforward process:

1. Implement a data infrastructure to obtain high-quality raw data.
2. Have the appropriate analytical capacities to distill data into a useable form.
3. Present the data in a format for optimal interpretation.

Although the framework itself is simple, the implementation is all for naught if the organization does not foster a DDDM culture. To achieve this culture, the organization must consistently base their decisions on data. As soon as organizations start to operate outside of the DDDM framework and base decisions on "gut feels" or "hunches," the culture will erode. Although it can be much easier to operate with a "Wild West" mentality, where supporting data is an afterthought at best, too many poor decisions will risk the organization being outpaced by the competition. Instead, with a DDDM culture, the organization becomes proactive rather than reactive, using data to drive decisions that improve systems, processes and strategies.

Summary

Implementing a DDDM framework can require significant investment of resources. However, these investments are necessary for any organization to evolve and begin to solve problems based on data-generated knowledge. By following the framework, your organization can catch the DDDM wave. However, without the corresponding DDDM culture, your organization might be left in the wake.

To assist with implementation of a DDDM framework by their clients, Bioclinica developed the Clean Patient Tracker (CPT), a Clinical Data Management tool, to obtain high-quality raw data by aggregating eSource data from multiple sources (EDC, ECG, Safety Labs, ePRO) into one comprehensive source database and to distill that raw data into relevant decision-making metrics. Clean Patient Optics (CP Optics) is a web-based visualization tool based on the aggregated data in the CPT, presenting the metrics in dashboards for interpretation and use by management teams.

Sources:

– *Bioclinica CP Optics White Paper*

– *Bioclinica eSource White Paper*

– *Report: A Conceptual Framework for Data Driven Decision Making*

– *http://whatis.techtarget.com/definition/data-driven-decision-management-DDDM*

– *https://www.fangraphs.com/library/misc/pitch-fx/*

– *https://econlearningstyles.wordpress.com/visual-and-verbal-learners/*