3S MEDIDATA

A (R)evolutionary Path
for Data Managers

The guide to driving data quality across the "5 Vs" of clinical data

TABLE OF CONTENTS

INTRODUCTION

The complexity of clinical trial data management has surged over the past decade.

Data managers are navigating unprecedented challenges in capturing, processing, and analyzing explosive volumes of data from an increasing variety of sources.

The new data landscape directly impacts their ability to ensure data quality, implement procedures to validate data integrity, manage resources efficiently, and control costs, especially in large-scale complex trials.

To stay ahead of the disruptive clinical trial data evolution, Data managers have an opportunity to adapt and adopt new approaches to managing and reviewing data efficiently.

Following up on "The 5Vs of Clinical Data" topic brief published by the Society for Clinical Data Management (SCDM) in March 2022[1], this eBook outlines how technology continues to inform and optimize clinical data strategies and management processes to effectively address the "5Vs"—Volume, Variety, Velocity, Veracity, and Value—of data that are reshaping clinical data management.

1. The Society for Clinical Data Management. (March 2022). "The 5Vs of Clinical Data".
   https://scdm.org/the-5vs-of-clinical-data/

VOLUME
MANAGING THE DATA DELUGE

Clinical trial data volume has increased seven-fold in the last 20 years to

3.6 million
data points

in a typical Phase III study.[1]

Over the past decade, clinical trial data volume has witnessed an unprecedented surge, growing tenfold. On average, Phase 3 studies now grapple with an overwhelming 3.6 million data points.

Compounded by staffing shortages, the explosion in data volume has far-reaching implications for data managers:

- How to collect, query, and manage the growing size of datasets?
- How to review, analyze, and extract meaningful insights for faster decision-making?
- How to address protocol changes without disruption?

Processes that once ensured efficiency are now strained, taking a toll on data managers affecting timelines, resource allocation, and the overall efficiency of studies.

Advanced technologies that critically improve data volume management include:

1. Average number of data points in Phase 3 Pivotal trials. 2001-2005 - 0.5 Million. 2015-2020 - 3.5Million.
   Source: Tufts CSDD 2021 (data is pre-pandemic).

Machine learning algorithms

Enhance data processing capabilities, by enabling systems to learn from historical data patterns.

Machine learning algorithms are designed to streamline and accelerate the identification of outliers, anomalies, and trends within massive datasets.

AI-powered predictive analytics

Identify potential data-related issues and allow proactive measures to be taken to reduce the likelihood of delays and errors.

Developing intelligent systems that automate the query management process significantly reduces the burden on data managers.

Flexible unified platform

Gain a scalable, secure infrastructure to handle large volumes of data.

Adopt a unified platform capable of ingesting, storing, managing, and analyzing all types of data, delivering a more agile and resource-efficient approach.

Leveraging technology to handle data volume drives efficiency in other critical clinical trial areas:

Implement real-time monitoring

The latest data management solutions enable immediate identification of data anomalies or discrepancies. This proactive approach allows timely intervention and correction, preventing potential setbacks during the clinical trial lifecycle.

Enable collaborative platforms

Facilitate seamless communication and data sharing among stakeholders to streamline the review and resolution process in real-time, fostering efficiency and transparency.

Invest in data quality management

Automate the identification and cleaning of inconsistent or inaccurate data, reducing the manual workload on data managers and enhancing the overall dataset quality.

Incorporate advanced data visualization tools

Allows stakeholders to interpret complex datasets quickly. Interactive dashboards and visual representations enable faster decision-making, reducing the time spent on manual data analysis and review.

Keep pace with the increasing data volume by integrating advanced technologies to streamline processes and provide a foundation for more efficient, accurate, and valuable clinical trials.

VARIETY
THE COMPOUNDING FACTOR

Compounding the complexity of sheer data volume, the variety of data sources in clinical trials continues to increase exponentially. Sources have become more diverse, from sensors, imaging, and labs, to Electronic Health/Medical Record (EHR/EMR) systems.

While rich in potential insights, data variety presents a formidable challenge for data managers. Understanding the intricacies of managing varied data sources is crucial in selecting solutions that seamlessly integrate and reconcile disparate datasets.

Aggregate and Standardize Data as it is Captured or Ingested

Understand the full context of the data as soon as it's available. Benefit from a source-agnostic view of the patient driving the activities required by all stakeholders, including data managers, and other stakeholders on the study.

### Use Multi-Source Data Acquisition and Import Technology

Site and patient data capture experiences for EDC, eCOA, Imaging, Sensors, and EHR data. Simple, powerful data ingestion that makes it easy to reliably import and validate.

Accelerate data quality

Data acquired across multiple sources is immediately available for further transformation, enrichment, and analysis.

Having the systematic mechanisms to acquire and aggregate the data, regardless of sources, optimizes insights and increases their ability to improve study outcomes.
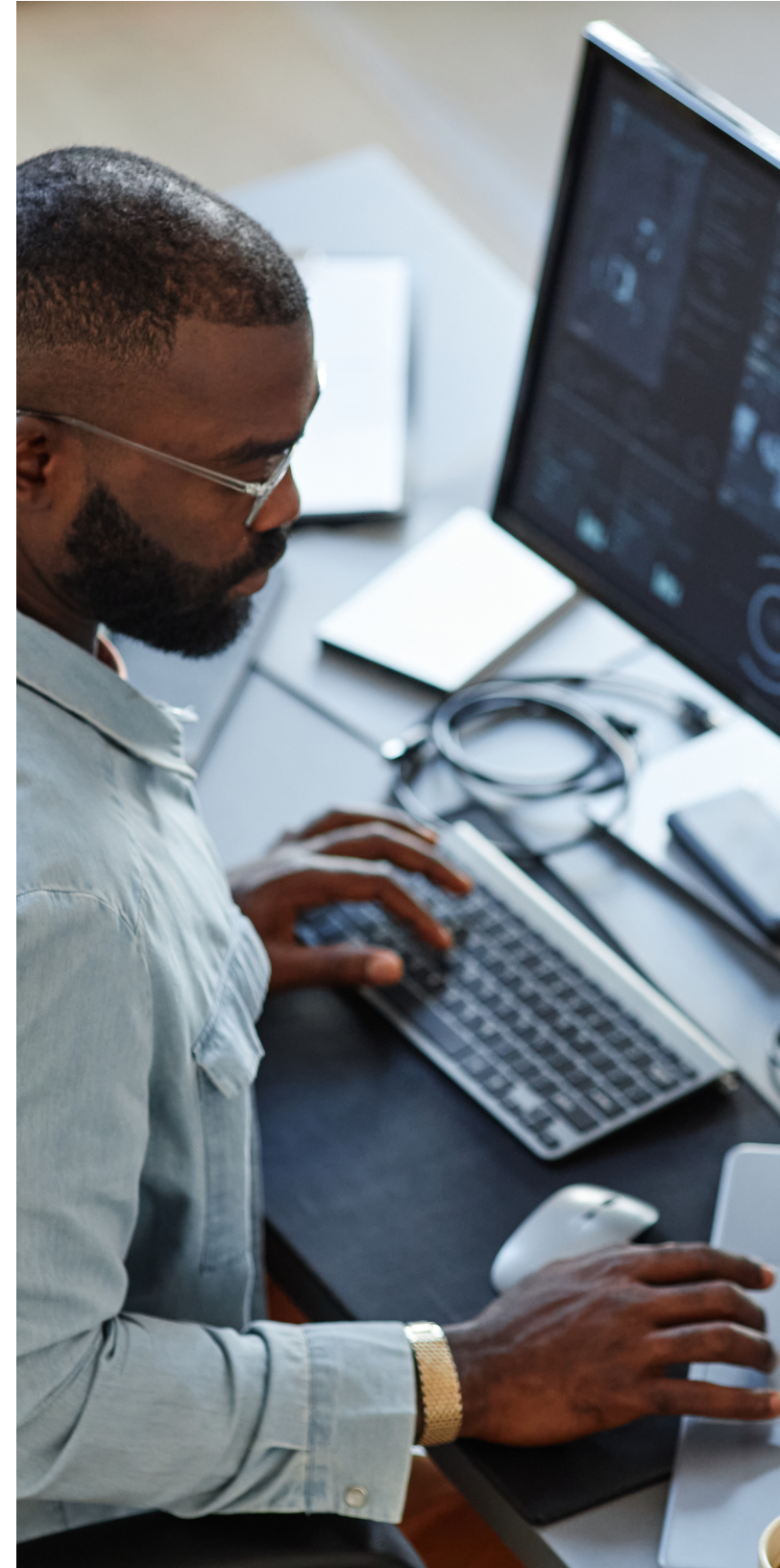
VELOCITY
ADAPTING TO THE REAL-TIME RUSH

While most trial data sources provide periodic data and traditional tools enable batch reviews, newer data sources, driven by sensors and apps, provide continuous data streams, altering trial procedures and timelines.

Real-time data requires rapid processing, validation, quality assurance, and immediate decision-making as insights and issues arise. Data managers must now proactively monitor and respond to trends and anomalies on the fly, rather than waiting for scheduled checkpoints.

This fundamental shift necessitates innovative tools.

Predictive Analytics

Use historical trial data and real-time information to forecast potential outcomes and trends. Proactively identify patterns, foresee challenges, and optimize decision-making by anticipating future developments in the trial.

Automated Anomaly Detection System

AI-driven algorithms analyze incoming data in real-time, automatically flagging anomalies or deviations from established patterns.

Risk-Based Monitoring System

AI-powered data assessment based on predefined risk factors. Prioritize monitoring efforts in real-time to allocate resources, concentrating on where deviations may pose the greatest risk.

Alert Systems and Notifications

Get real-time alerts and notifications when predefined thresholds or deviations are surpassed, triggering quick corrective actions to maintain data integrity.

These tools enable data managers to derive valuable insights, boosting the efficiency and effectiveness of clinical trials.

VERACITY
DATA FOR REGULATORY SUCCESS

Data accuracy and consistency are crucial for successful regulatory submissions, to meet requirements for accuracy, consistency, and completeness. Error-free, consistent data speeds up their review process and reinforces the credibility of the study results.

Veracity confirms data is accurate, complete, consistent, and aligned with predefined attribution and conformity standards. Integrating advanced analytics into data quality control processes ensures data is authenticated, interpreted, and safe.

These tools analyze data and patterns to:

- Ensure data quality and consistency across diverse data sources for adherence to data quality targets.
- Maintain interoperability standards to ensure the veracity of integrated datasets
- Establish robust data governance frameworks, defining roles, responsibilities, and processes

To extract optimal value from data, veracity is key to:

- Intelligent trial design, improving protocol development, patient selection, and endpoint optimization.
- Using synthetic control arms and accurate virtual patient cohorts, streamlining processes, and optimizing resources.
- Training machine learning algorithms, enhancing data analysis, and providing insights into patient responses and risks.

VALUE
MINING EACH DATA POINT

The value of clinical trial data goes far beyond its immediate use in clinical trials. The true value of data lies in all the applications and long-term benefits that can improve the impact of clinical trials.

Leverage Synthetic Control Arms (SCA)

Utilize existing data, from historical trials, to simulate a control group. Synthetic control arms speed up the trial process, reduce costs, and minimize the need for patient recruitment.

Optimize Resource and Patient Recruitment

Using SCA drives trial resource efficiencies, by reducing the logistical burden and cost associated with large-scale patient recruitment. It ensures that patients enrolled receive beneficial treatments, rather than placebos or standard care.

Power Intelligent Trial Design

Clean data, continuously training Machine Learning algorithms, enhances data analysis capabilities, delivering insights into patient populations, treatment responses, and potential risk factors.

Personalized Medicine

Analyzing detailed data from diverse patient groups fosters more personalized treatment approaches, tailored to individual patient needs and genetic profiles.

Drive Regulatory Compliance and Approval

Data accurately analyzed and interpreted can provide strong evidence for the safety and efficacy of a treatment, which is essential for regulatory approval.

Data provides invaluable insights into disease patterns, treatment responses, and patient demographics, which can inform future trials and medical research.

CONCLUSION

Traditional processes no longer scale. The transformation of clinical data management takes a different approach to keeping up with the new trial landscape.

Medidata's suite of solutions, including machine learning algorithms, predictive analytics, and cloud-based infrastructure, is essential in designing a successful clinical data strategy to manage its exponential growth and complexity.

The shift to real-time data requires adaptable strategies, and Medidata's technology is instrumental in managing the complexities of continuous and heterogeneous data.

Medidata ensures data quality with real-time validation tools and advanced analytics, delivering data beyond regulatory compliance. It drives numerous aspects of trial management, from synthetic control arms to intelligent trial design and patient insights.

Medidata's tools and insights are vital for data management professionals to effectively manage the 'Five Vs' complexities and extract the full value from clinical trial data.

As you navigate the evolving world of clinical trials, Medidata is the trusted technology for achieving greater trial outcomes and faster innovation.

MANAGING THE "FIVE Vs" OF CLINICAL TRIAL DATA

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 2,200+ 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™. Medidata, Medidata Rave and Acorn AI are registered trademarks of Medidata Solutions, Inc., a wholly owned subsidiary of Dassault Systèmes.

For more information: **contact-sales@medidata.com** | **+1 866 515 6044**

3S MEDIDATA