MEDIDATA | WHITE PAPER

Back to Basics: A Survey of Clinical Data Management System Users and Decision Makers

Table of Contents

Introduction

Modern clinical trials continue to increase in complexity, with an expanding array of high-volume and high-velocity data coming from a wide array of sources, including decentralized clinical trial (DCTs), adaptive designs, use of real-world evidence, sensors, and more.

A recent estimate indicates that data points collected in Phase III trials have increased threefold in the last 10 years, reaching an average of 3.6 million (Tufts CSDD, 2021). This trend will undoubtedly continue, with 70% of trials expected to incorporate digital health technologies (DHTs)—including high-volume and high-velocity wearables and sensors—in the next few years (Myshko, 2019).

Many of these data streams now come from sources outside of electronic case report forms (eCRFs) and are increasingly managed outside of electronic data capture (EDC) systems (Zozus, 2021; Wilkinson, 2019), which further compounds the issue trial data silos.

This has created significant data management challenges, as traditional clinical data management technologies and manual processes to review, clean, and lock data have not progressed as fast. Data is often managed with tools and processes that are not able to work at the speed required to support the realities of today's clinical trials. As clinical trials become more digital, sponsors and contract research organizations (CROs) are turning towards Clinical Data Management Systems (CDMS) to collect, validate, and transform their data across the lifecycle of a trial, including study design, planning, conduct, closeout, and post-trial analysis.

Given this context, Medidata sought to better understand how small, medium, and large biopharmaceutical companies are managing their data in this complex environment and gain their perspectives on what they are looking for in a CDMS that handles study design, planning, execution, analysis, and submission. Medidata commissioned a survey of 102 buyers and users of clinical data management tools, including managers through C-suite executives, and spanning a wide range of roles, including clinical operations, data management, data science, therapeutic area heads, and information technology (IT).

The survey asked participants about their current pain points, current CDMS solutions, and features/functionalities they desire in a future CDMS. While the study identified a few notable differences between large and small to medium companies, for the most part, all companies, no matter their size, are facing similar challenges and seeking similar solutions. The overarching conclusion from the data is that buyers and users have big pain points with data integration, largely driven by increasing trial complexity and the concurrent surge in data volume. Respondents also indicated that they are not yet seeking more advanced features, such as search, semantic layers, and metadata management.

This white paper discusses the key results of this survey.

Key Takeaways

Data Integration and Reconciliation:
Approximately 55% of respondents indicated that integrating and reconciling data from various sources is their greatest pain point. Though this is not a new finding, current approaches to data integration are not meeting most customer needs. With the aggressive adoption of DCT trial technologies and services over the past few years, which is expected to continue (Agrawal, 2021), it is reasonable to assume that this issue will be exacerbated if modern data management tools are not embraced and deployed.

Buyer and User Needs:
Data integration and core data management functionality were ranked much higher than more advanced features such as a semantic layer, expanded, and metadata management. Independent of company size, buyers and users identified large pain points with respect to basic data management needs but are not yet ready for more innovative features.

Data Access:
Respondents were split on how they wanted to access trial data, with 41% seeking seamless access within a provider's CDMS environment and 39% preferring access within a common statistical environment, such as SAS or R. The least preferred method was an in-house statistical environment.

Historical Data Integration:
Respondents were largely split on their current methods for integrating historical clinical trial data, with one-third doing so outside of a CDMS and another one-fifth not currently integrating it at all.

At a high level, all respondents, independent of company size, would like the following features in a seamless end-to-end CDMS:

- **Core Data Management**
- **Data Ingestion and Integration**
- **Edit Checks**
- **Issue Management**
- **Study Design**
- **Data Visualization and Exploration**
- **Security**

The following features tended to be ranked lower:

- **Operational and Financial Performance Reporting**
- **Advanced Functionality Search**
- **Semantic Layer**
- **Metadata and Aata Model Management**

Methods and Definitions

Medidata Solutions sponsored a blinded, web-based, quantitative survey. A third-party market survey company recruited the experts using a proprietary panel and surveyed in February 2022. Participants were provided an honorarium for their time. Participants were screened to ensure they had the appropriate level of decision-making experience. The recruitment criteria were as follows:

- Currently employed by a pharmaceutical, biopharma, or biotech company or a contract research organization (CRO) - Company size was defined by annual revenue as **large** (greater than $1B), **medium** ($300M–1B), and **small** (less than $300M)
- Titled as manager, director, president/SVP/VP, or C-level executive

- Functionally serving as clinical operations/project manager/CTL/asset manager; data management; data scientist/biostatistics; digital innovation; executive corporate management; general IT; clinical trial technology IT; medical/therapeutic area head; outsourcing/procurement management; or R&D management
- Moderately involved in data management and acting as a final decision maker, part of a team that makes decisions, or an end-user of externally developed technology products/solutions for data management
- From a company with assets beyond Phase I (Phase II-IV)
- Moderately familiar with the CDMS technology concept

The number of participants for each category was as follows: large (n = 56), medium (n = 23), and small (n = 23). Overall, 90% of the respondents were from pharmaceutical, biopharma, or biotech companies and 10% from CROs, with 78% based in North America and 22% in Europe.

Statistical testing was performed at a 95% confidence interval to identify differences among respondents across company sizes. Segments required at least 30 respondents for statistical testing.

Table 1 provides definitions for commonly used terms that respondents were asked to refer to throughout the survey.

Table 1: Explanations of Commonly Used Terms Associated with CDMS for This Survey.

Explanations of Commonly Used Terms	
CDMS (Clinical Data Management System)	Supports seamless end-to-end CDM processes, including study design, execution, analysis, and submission. At a high level, the CDMS would allow users to: • Design the data collection process, including centralized study design and metadata management • Ingest data from multiple sources and transform it to a common standard with clear traceability • Access data—through either streams or analytics and visualization tools—in real-time, with a single administrative console • Apply advanced data science capabilities—including artificial intelligence and machine learning—to clean and transform data, assess risks, and expose hidden trends
Study Design	• Efficiently and accurately design the data collection and data structures to reflect the study protocol • Build and validate ways of ingesting these data (e.g., EDC and ePRO forms, sensors) • Maintain a template library (global library) of reusable components for an organization, including industry-standard and customer-specific templates. • Configure edit checks and data transformations • Support collaborative study design process and approval and release workflows
Metadata and Data Model Management	• Assign metadata to discover, extract, store, manage, index, and catalog objects • Allow metadata-driven transformation into required standards, input files, or harmonized data models • Define and manage reference and master data • Maintain metadata, contextualize, and manage between versions throughout the study • Automate processes and rule creation/application based on metadata

Explanations of Commonly Used Terms	
Data Ingestion and Integration	• Ingest real-time clinical trial data, including wearables and sensors • Integrate APIs, and interact with data from multiple sources, including historical trial data • Ingest and normalize dissimilar data sets from any internal or external source
Data Management—Core	• Efficiently manage complex data mapping • Aggregate data sets and create and manage snapshots, data segmentation, and data archival • Centralize coding
Data Management—Edit Checks	• Run data quality checks and reconciliation • Audit logs of data and edit check resolution • Run machine-learning-based edit checks
Data Management—Issue Management	• Query management across all data
Semantic Layer	• Associate study data with common concepts and vocabularies • Achieve a shared understanding of data across therapeutic areas, studies, phases, and events
Advanced Search Functionality	• Offer full-text search across data
Data Visualization and Exploration	• Have a visual and graphical way of collating and interacting with data sets • Apply analytic tools to identify anomalies • Conduct medical reporting and produce information and reports • Conduct performance and trend analysis on past and current studies
Operational and Financial Performance Reporting	• Integrate with payments and budgeting systems • Track spending vs. budgets for sites, CROs, and other vendors
Data Access	• Provide user-friendly, real-time, centralized data access
Security	• Offer single sign-on and multifactor authentication • Provide role-based permission data access • Encrypt data in motion and at rest

Findings

This section provides a summary of the results from the survey's key focus areas:

Current Management Practices and Pain Points | **Relative Importance of CDMS Features** | **Overall Reactions to the CDMS**

Current Data Management Practices and Pain Points

The following are questions posed to respondents:

1. What types of solutions are you currently using to meet the needs that this CDMS would aim to fulfill?
2. Which of the below methods do you currently use to integrate historical data?
3. Thinking about your current system, please describe the biggest "pain points" you encounter in managing clinical data.

High Level Takeaways: No Single Solution Currently Meets All Respondent Needs

Most respondents indicated that they are currently using several solutions, including a mix of in-house and externally developed solutions, to fulfill the needs satisfied by the CDMS described in the survey. Generally, more respondents cited third-party solutions instead of "homegrown" ones. Collectively, these data highlight that currently no single solution can satisfy all the respondents' needs. Overall, roughly 40% of respondents indicated they use a CDMS in some capacity via either single (18%) or multiple (23%) data models.

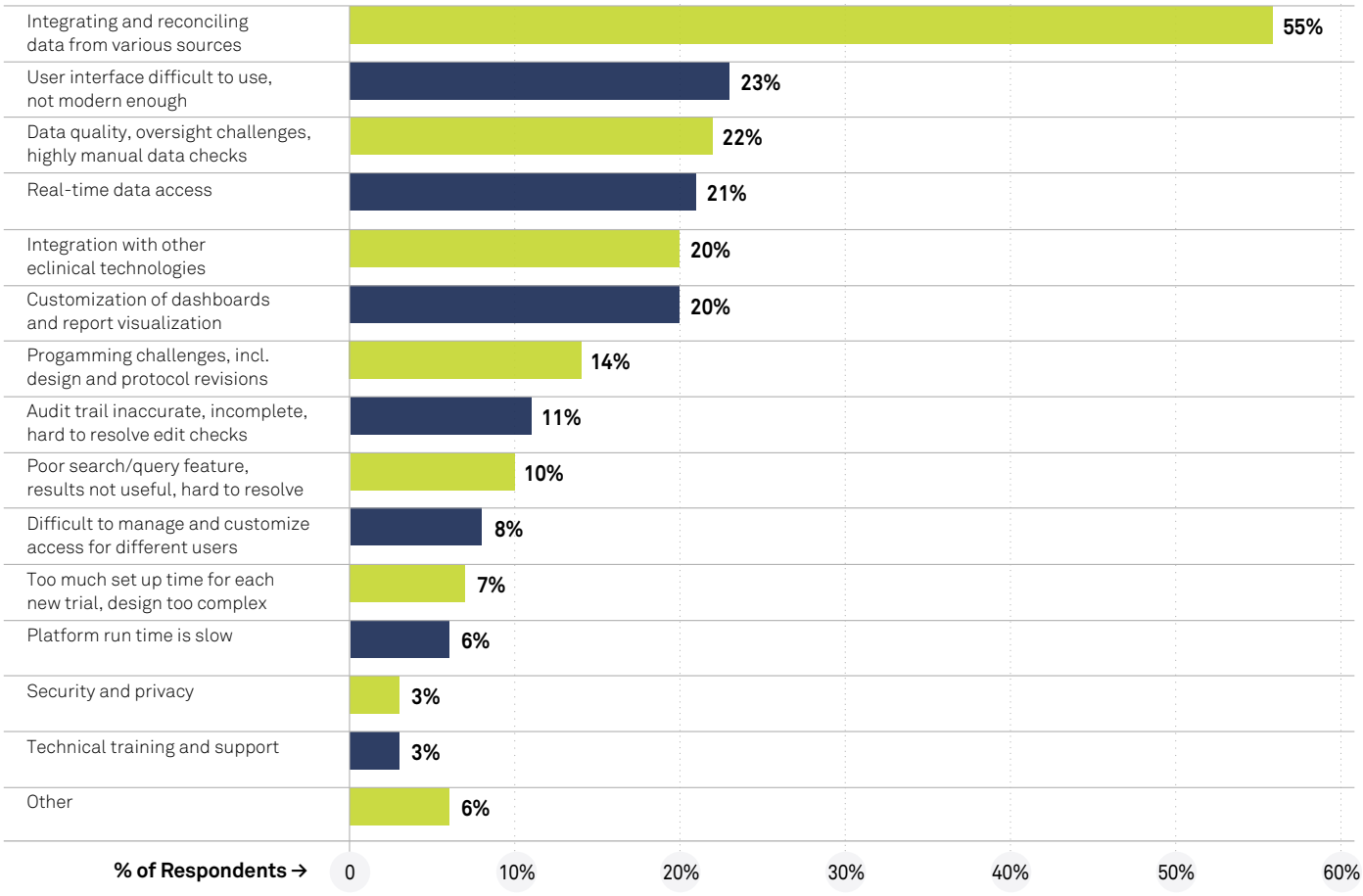
Historical Data Integration Processes Vary

Respondents were split on their methods for integrating historical data, with 41% indicating they integrate historical data within a CDMS, 33% indicating they do so outside of a CDMS. A further 20% of respondents indicated that they do not integrate historical data.

Data Integration and Reconciliation Pain Points Persist

Respondents described several pain points related to how they manage clinical data. The most mentioned challenges included ingesting, integrating, and aggregating data from various sources. Other common themes included complaints about the user interface, data quality and access issues, integration with other technologies, customization of reports, and data visualization.

Figure 1: Pain Points in Clinical Data Management



Importance of CDMS Features

High-Level Takeaways: Data Integration/Reconciliation and User-Friendly Interface Are Most Important

In alignment with respondent pain points, respondents indicated (in an unaided/unprompted manner) that data integration/reconciliation and a user-friendly interface as the most essential features a CDMS should have.

Security Was Identified as a "Must-Have" by 75% of Respondents

Security, the feature described as single sign-on, multifactor authentication, role-based permissions, and data encryption, was considered a "must-have" by three-quarters of respondents. Following security, ~60% of respondents believe "study design, data access, and data management—edit checks" to be "must-haves."

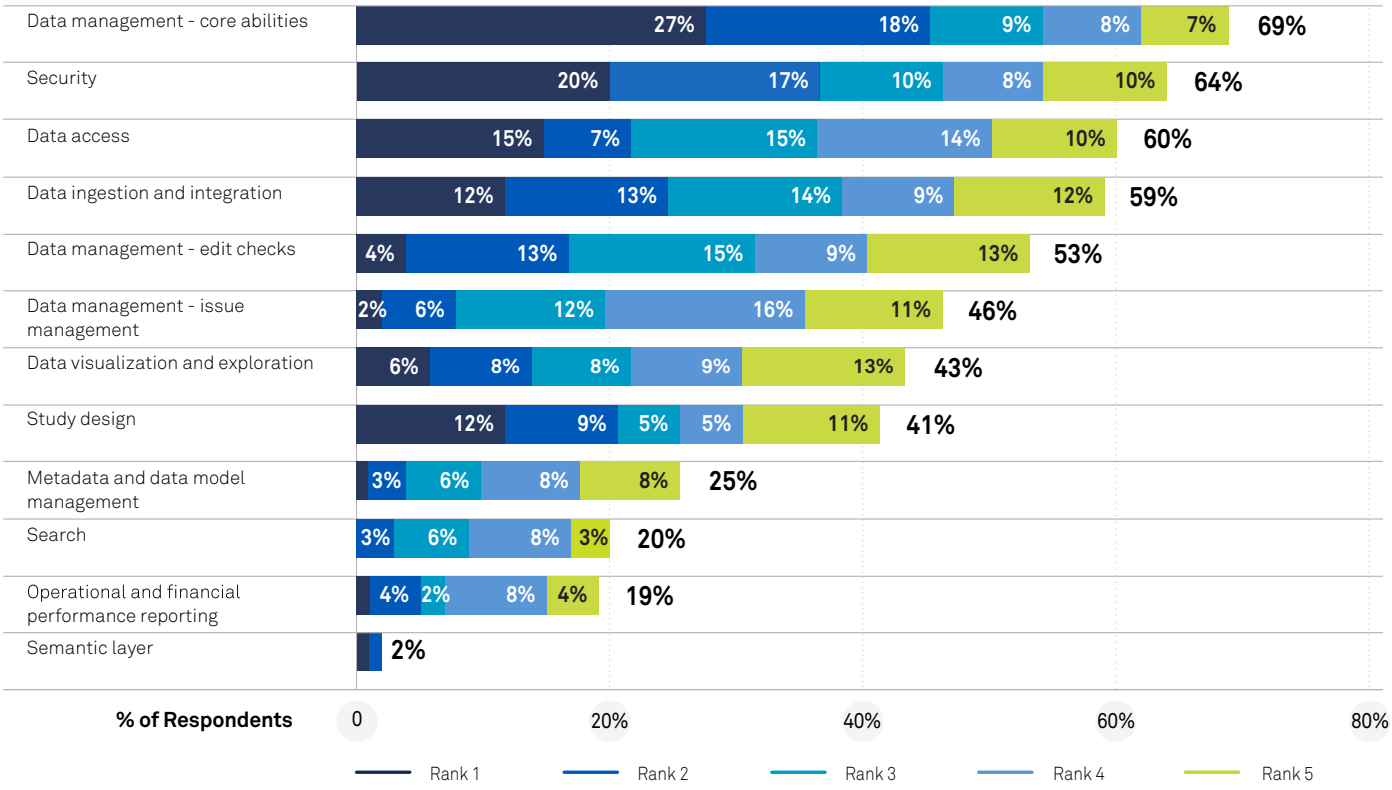
Notably, a significantly higher proportion of respondents from large companies (82%) selected "data ingestion and integration" as a "must-have" or "should have" compared to respondents from other companies (61%).

> **"Ability of the system to integrate data from multiple points/systems into one, user-friendly, easy to access system to enable streamlined decision-making."**
> **Survey Respondent**

Core Data Management Ability Was Ranked as the Most Important CDMS Feature

Data management—core abilities (management of data mapping, aggregated data sets, managed snapshots, data segmentation, archival, and centralized coding) were ranked as the most important feature of a CDMS. Overall, according to the highest percentage of respondents, the top five features in terms of importance were data management—core abilities, security, data access, data integration, and ingestion, and edit checks.

Figure 2: Ranking Importance of CDMS Features.



Preferred Data Access Method

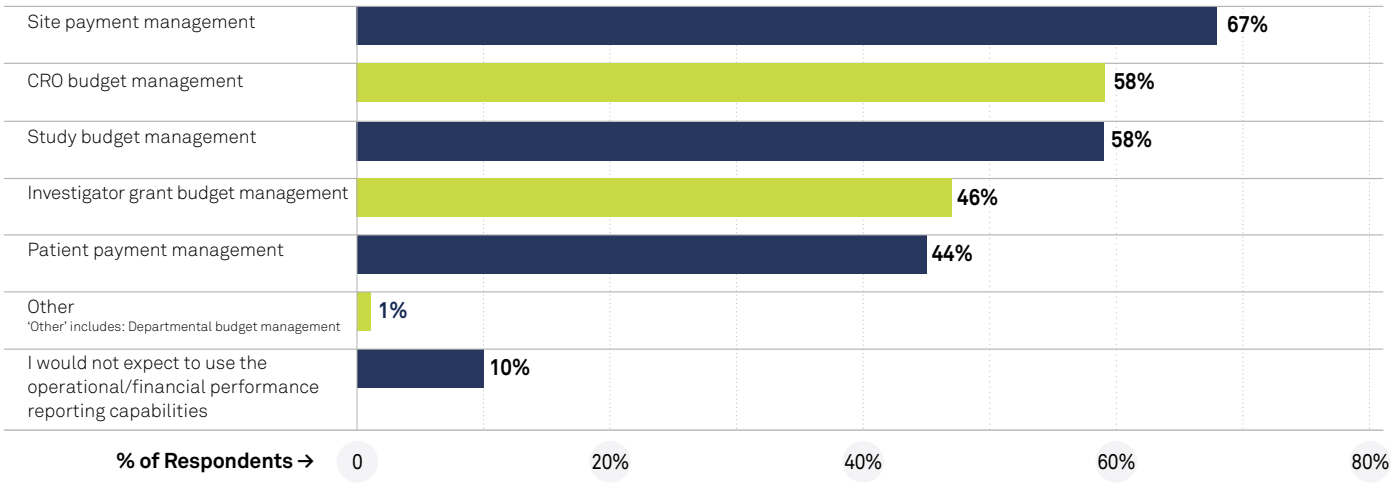
Forty-one percent (41%) of respondents preferred accessing data within the CDMS solution, which was a similar proportion to those preferring a common statistical environment, such as SAS or R. An in-house statistical environment was the least preferred method (20%).

> **"Most important features of any CDMS (not per se in order of priority) are that they should be user-friendly (easy and self-explanatory to use), able to easily and comprehensively capture data from patients, and clinicians, and nowadays also from smart technology (devices, etc.). It should allow for quick and easy setup of trials without minimum training requirements, should allow for support in user friendly ways (e.g., virtual calls via platforms like Zoom, etc.), allow for efficient running of the system and creation of any reporting requirements, export of data and related analytics, and support in ensuring readiness for filing in the most convenient manner."**
>
> **Survey Respondent**

Operational/Financial Reporting Types

Almost 70% of respondents selected "site payment management" as a type of preferred operational/financial performance reporting, while nearly 60% selected "CRO budget management" or "study budget management." Ten percent (10%) indicated that they would not expect to use this type of reporting within a CDMS.

Figure 3: Operational/Financial Reporting Types



Data Visualization/Exploration Feature Importance

Three data visualization and exploration features were nearly tied in terms of their relative importance: "conduct medical reporting and produce information and reports" (29%), visual and graphical way of collating and interacting with data sets (29%), and "analytical tools to identify anomalies." The highest proportions of the Top 2 rankings were "conduct medical reporting and produce information and reports" and "analytical tools to identify anomalies received."

Metadata/Data Model Management Feature Importance

Respondents were generally split on which metadata/data model management features would be most important in a CDMS; 22–25% ranked these features as the most important:

- "Assign metadata to discover, extract, store, manage, index, and catalog objects";
- "Allow metadata-driven transformation into required standards, input files, or harmonized data model"; and
- "Define and manage reference and master data."

Semantics and Search Feature Importance

Forty-three percent (43%) of respondents ranked "full-text search across data" as the most critical semantics and search feature. This was followed by "a shared understanding of data across therapeutic areas, studies, phases, and events," with 37% of respondents ranking this as the most important.

Customer Reactions to Proposed CDMS

The following are the three questions posed to respondents:

1. Overall, how compelling do you find this CDMS for filling all your data management needs?
2. How believable would you find the idea that your company could fill all its data management needs using a single CDMS?
3. If your company were considering purchasing a CDMS, what would you consider the greatest barriers to adoption?

High-Level Takeaways: Appeal of CDMS for Fulfilling all Data Management Needs

Overall, the concept of the CDMS resonated well with respondents. Forty-one percent (41%) found the concept of the CDMS to be "extremely compelling," followed by 51% who considered it to be "moderately compelling." Few respondents thought the CDMS to be only "slightly compelling" or "not at all compelling."

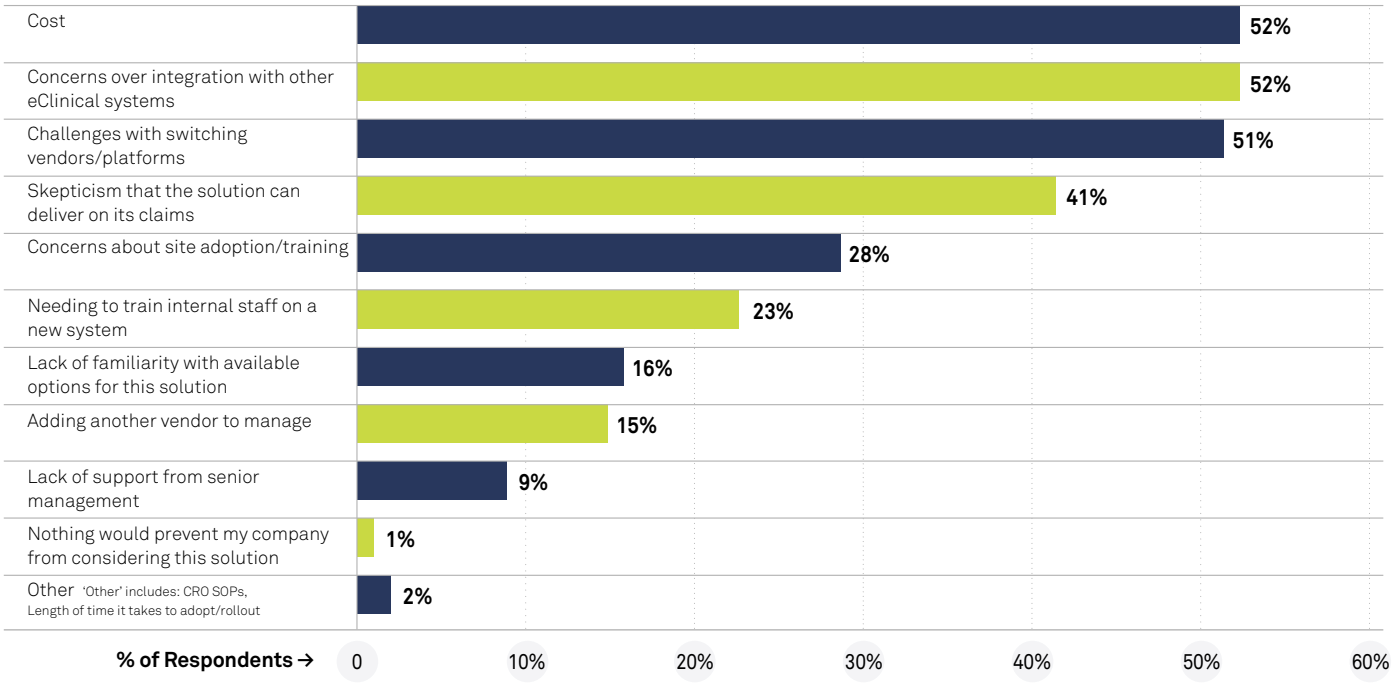
Believability of a Single CDMS to Satisfy All Data Management Needs

Respondents were less convinced that the proposed CDMS solution could fill all of their company's data management needs. Most respondents found this to be "moderately" or "extremely" believable, but almost 30% considered it to be only "slightly" or "not at all" believable. Respondents at small and medium companies (83%) found this to be significantly more believable than those at large companies (63%).

Barriers to CDMS Adoption

About half of respondents selected three barriers to CDMS adoption: cost, concerns over integration with other e-clinical systems, and challenges with switching vendors/platforms.

Figure 4: Barriers to CDMS Adoption.



Summary

The results of this survey confirm that in today's increasingly complex clinical trial landscape, with exponential data growth, the industry is looking for Clinical Data Management Systems (CDMS) that not only provide core data management capabilities—such as management of data mapping, aggregated data sets, managed snapshots, data segmentation, archival, and centralized coding—but also overcome major pain points related to the organization and integration of this data.

This is because traditional and manual methods of data aggregation, reconciliation, and transformation are not adequate to deal with the massive clinical volumes of data produced by modern trials, which regularly incorporate multiple data standards from many different digital tools. This problem is further exacerbated by the fact that many data streams are now coming from sources that are managed outside of primary EDC systems. Consequently, trial data cannot be quickly and efficiently collected, connected, and analyzed since the data are housed in silos and not integrated within a CDMS.

Furthermore, the COVID-19 pandemic accelerated the industry's adoption of Digital Health Technologies (DHTs), and this trend is expected to continue in this increasingly digital environment. We expect that the pharmaceutical industry will evolve and adapt to this new paradigm and increasingly move from single and fragmented point solutions to CDMS that are part of a unified platform. This approach will allow for seamless collection and management of data in a single platform, rather than integrating data from multiple sources and systems, eliminating /mitigating a major pain point.

This will ease the burden on clinical development teams and is scalable to support large numbers of studies without having to reinvent the wheel for every new study. It will allow sponsors to centralize their data so they can gain powerful insights across trials instead of having to deal with disparate data sets that are isolated across several point solutions. It will allow sponsors to collect, clean, and analyze their data sooner to bring life-changing treatments to patients faster.

For more information on Medidata Solutions, visit **www.medidata.com**.

References

1. Agrawal G, Moss R. et al. No place like home? Stepping up the decentralization of clinical trials. McKinsey & Company. June 2021. Available at: https://www.mckinsey.com/industries/life-sciences/our-insights/no-place-like-home-stepping-up-the-decentralization-of-clinical-trials

2. Myshko, D. Wearables in Clinical Trials. PharmaVoice.com. March 2019. Available at: https://www.pharmavoice.com/article/2019-03-wearables/

3. Tufts Center for the Study of Drug Development. 2021. January/February Tufts CSDD Impact Report: Rising Protocol Design Complexity is Driving Rapid Growth in Clinical Trial Data Volume. January/February Vol. 23 No. 1.

4. Wilkinson M, Young R, Harper B, Machion B, Getz K. Baseline Assessment of the Evolving 2017 eClinical Landscape. Ther Innov Regul Sci. 2019; 53(1): 71–80.