

CLINICAL DATA INDUSTRY RESEARCH

Two-Thirds of Data Managers and CRAs Warn of Data Quality Risks if Trial Inefficiencies Persist



Contents

1. Executive Summary	3
2. Key Findings	5
2.1 Inefficiencies faced by data managers	5
2.2 Inefficiencies faced by CRAs	7
2.3 Cost of queries	9
3. Root Causes of Unproductive Effort	9
3.1 Top drivers of inefficiency for data managers and CRAs	9
3.2 Tools and processes as sources of inefficiency	10
4. Implications of Lost Productivity	11
4.1 Risks and consequences of inaction for data managers and CRAs	11
4.2 Barriers to a more productive future	13
5. The Future Role of Clinical Data Professionals	14
5.1 Evolution of the data manager role	14
5.2 Evolution of the CRA role	14
6. Conclusion and Recommendations	15
7. Appendix	16
7.1 Methodology	16
7.2 Survey demographics	17
7.3 Full quantitative survey results	18
8. Acknowledgments	23

1. Executive Summary

Pharma and biotech teams face pressure to deliver faster trial timelines under increasingly complex protocols and constrained budgets. Meeting these demands requires not just innovation, but a willingness to depart from 'good enough' processes and technology for those carrying out clinical data tasks.

Inefficiency is the norm for many clinical professionals. Data managers and clinical research associates (CRAs) – the people responsible for collecting, monitoring, and ensuring the data is fit for purpose – are bogged down by manual processes and fragmented systems. Too much time is lost to repetitive steps, re-entry of information, and workarounds outside core clinical platforms.

This research identifies the scale and root causes of that lost productivity in typical Phase III trials, and offers a path forward. We surveyed 88 data managers and CRAs and conducted in-depth interviews with select participants to gather first-hand accounts of the impact of workflow inefficiencies.



69%

of data managers and CRAs worry about **data quality** as a result of inefficiency



81%

agree that better **system connectivity** would improve trial execution



12 hours

each data manager's time per week per study **consumed by the two most inefficient tasks**

Key Findings

I Inefficiencies are embedded in workflows

Most data manager tasks take more time than they should, especially manual data reconciliation and data review & cleaning. CRAs say updating multiple systems, writing reports, and following up with sites all take longer than they should.

I Manual work and disconnected systems are the root cause of inefficiencies

CRAs are impacted the most by system fragmentation, with data managers citing too many manual steps and a lack of modern tool usage as the main reason for inefficiency. Most manual data reconciliation work is happening outside of clinical systems, or using multiple different systems.

I The cost of inaction is burnout and data quality

Burnout or staff turnover and data quality issues were some of the greatest consequences for CRAs and data managers if inefficiencies are left unaddressed. CRA burnout and turnover is a widely reported issue, but could be an increasing trend for data managers.

I Organizational resistance to change hinders progress

Despite the challenges caused by inefficient processes and systems, almost half of data managers and CRAs say that resistance to change is a key barrier to a more productive future state. A sense of current workarounds being 'good enough' hides the damage caused by inefficiency in clinical trials. Nearly half of respondents think that SOPs fail to reflect real-world workflows or underutilize available tools, demonstrating the organizational barriers to greater efficiency.

I Future roles focus on pragmatic change

AI support and risk-based approaches are anticipated to evolve data managers' work over the next two years, but data managers cited automation as #1 for their role evolution. The vast majority of CRAs expect risk-based monitoring to become the norm in their roles, but feel the stretch from greater study complexity and budget/resource constraints.

I System modernization is a clear opportunity

Three quarters of data managers and over half of CRAs say their teams are actively looking for ways to modernize trial execution. Nearly all respondents feel that better integration between clinical data and operations systems would significantly improve execution.



Identifying inconsistencies between two systems for the same kind of data should be easy in the advanced tech world we live in."

Data manager's response to the question "Why is [manual data reconciliation] so inefficient?"

2. Key Findings

2.1 Inefficiencies faced by data managers

Data managers spend the greatest proportion of their time on data review and cleaning (18%) [Figure 1]. This task is also a top inefficiency, cited as the second most inefficient task (35%) after manual data reconciliation (52%). Manual reconciliation occupies about 11% of data manager's time yet is their highest priority to improve, indicating that inefficiency is a bigger driver of frustration than time spent on a task.



*We have home-grown tools for data cleaning but I don't trust them because they cannot handle the variability in data structure between studies. Data sources have differences in date formatting, extra spaces in an output, column naming... They can differ by source and by study. **There is too much variation so we stick with Excel.***

Senior Clinical Data Specialist,
Global CRO

FIGURE 1A

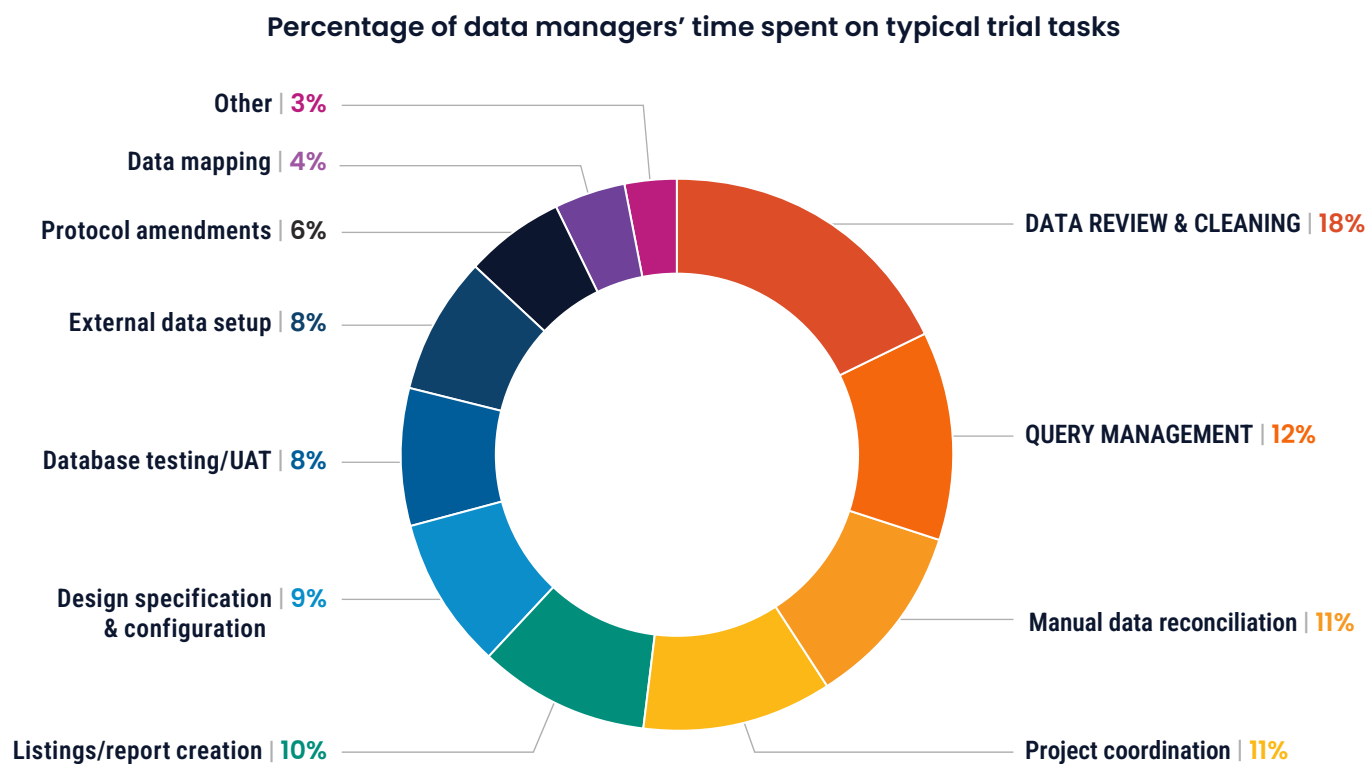
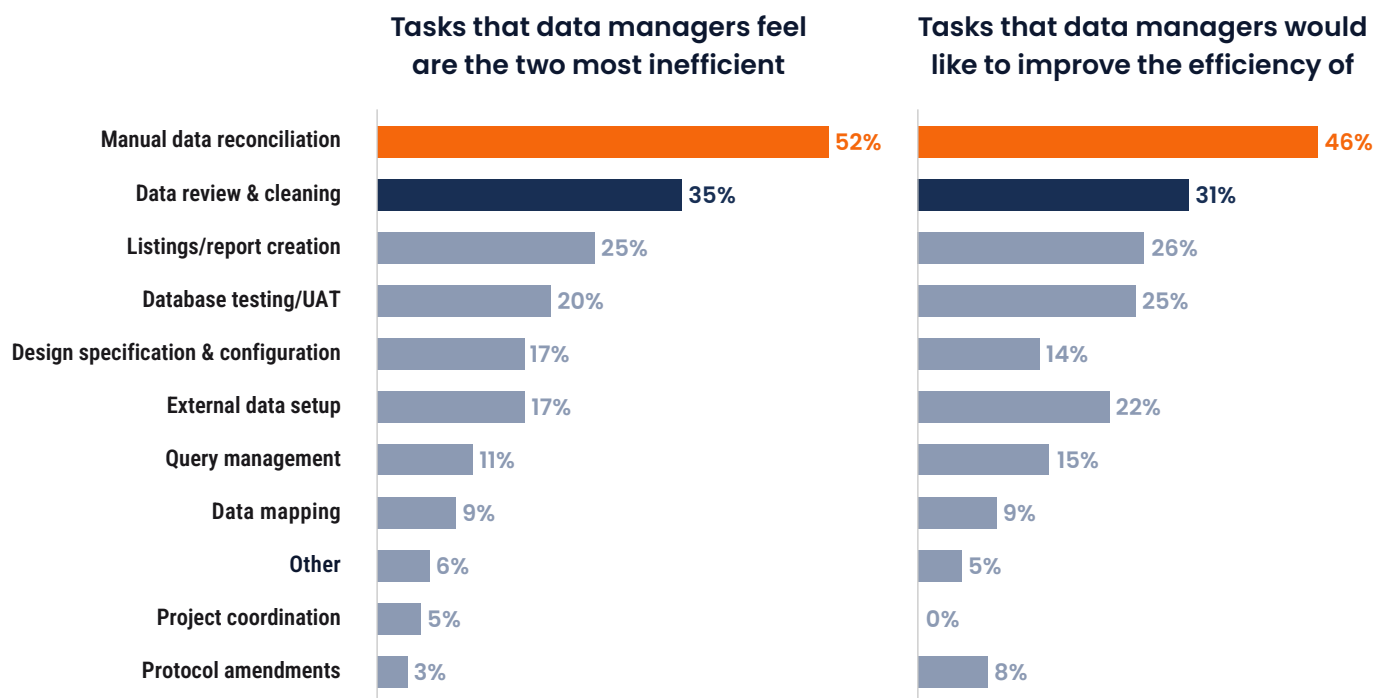


FIGURE 1B



Data captured via online survey from 65 data managers working on Phase III trials.

Tasks flagged as the most inefficient (manual data reconciliation and data review and cleaning) are also among the tasks more often performed outside clinical systems. For those citing data reconciliation as a top inefficiency, 97% perform reconciliations either outside of clinical systems (50%) or with a mix of systems (47%). For data review and cleaning, 39% of respondents perform the task outside clinical systems, 52% with a mix of systems, and only 9% within a single clinical system.

“

It feels like for every study, we're reinventing what data review is. There's a lack of good, reusable data cleaning programming, which would help with both [data cleaning and review and manual data reconciliation].”

Data Analyst, Small Biopharma

One senior data manager from a global CRO estimates that of the three hours each data manager spends on reviewing data, one hour is wasted reviewing unchanged data, and another hour is spent applying Excel formulas. According to this account, data managers are only spending one third of their data review time reviewing new data.

The senior data manager explains the challenges around a lack of insight into data changes during data review: *"I complete a round of reconciliation and the next month when I download the data there is no way to tell what data has changed. We apply a lot of Excel formulas and effort to figure out what's been updated and what queries got opened. Every month I pull the same reports and apply the same effort, even though 70% of the data is the same."*

2.2 Inefficiencies faced by CRAs

While CRAs spend most of their time on monitoring visits (22%), perceived inefficiencies center on documentation and manual tracking – tasks that collectively consume around 18% of CRAs' time yet generate the highest frustration [Figure 2]. Documentation and manual tracking are identified as top priorities for improvement by 44% of respondents.

FIGURE 2A

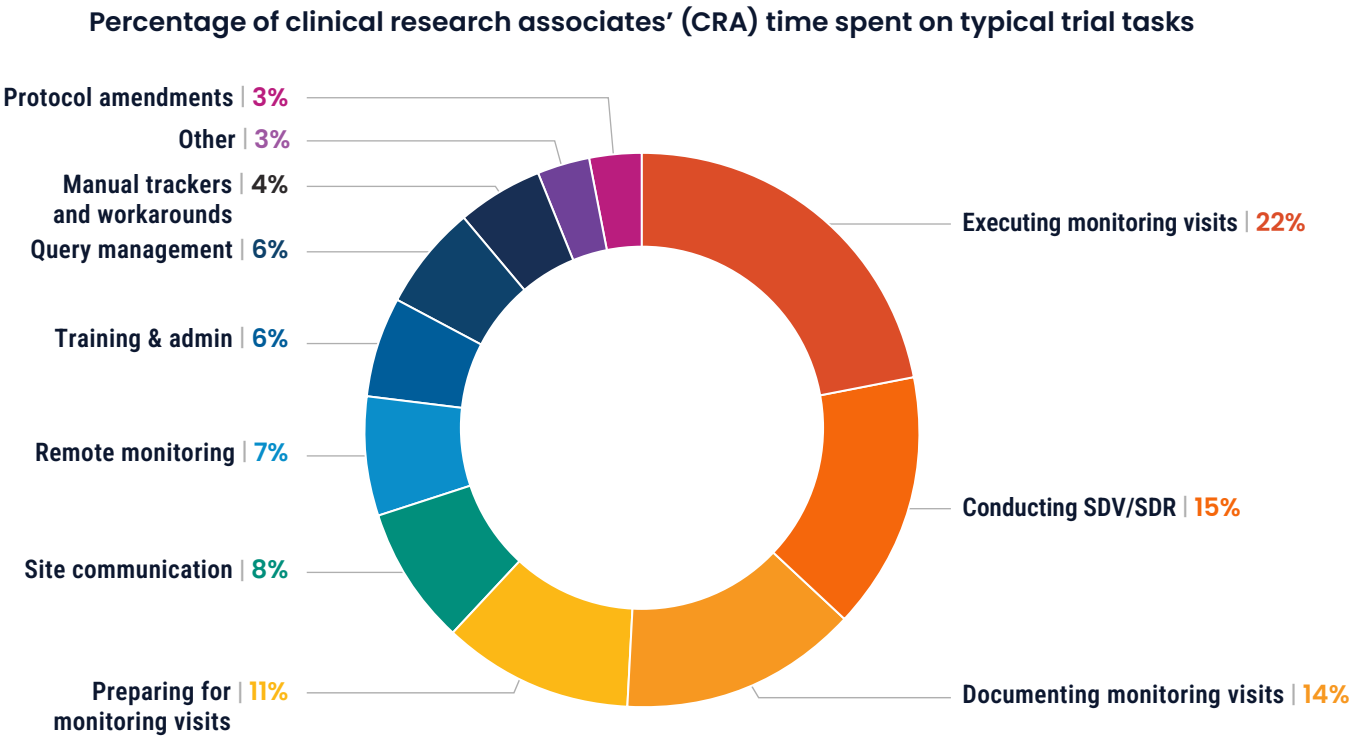
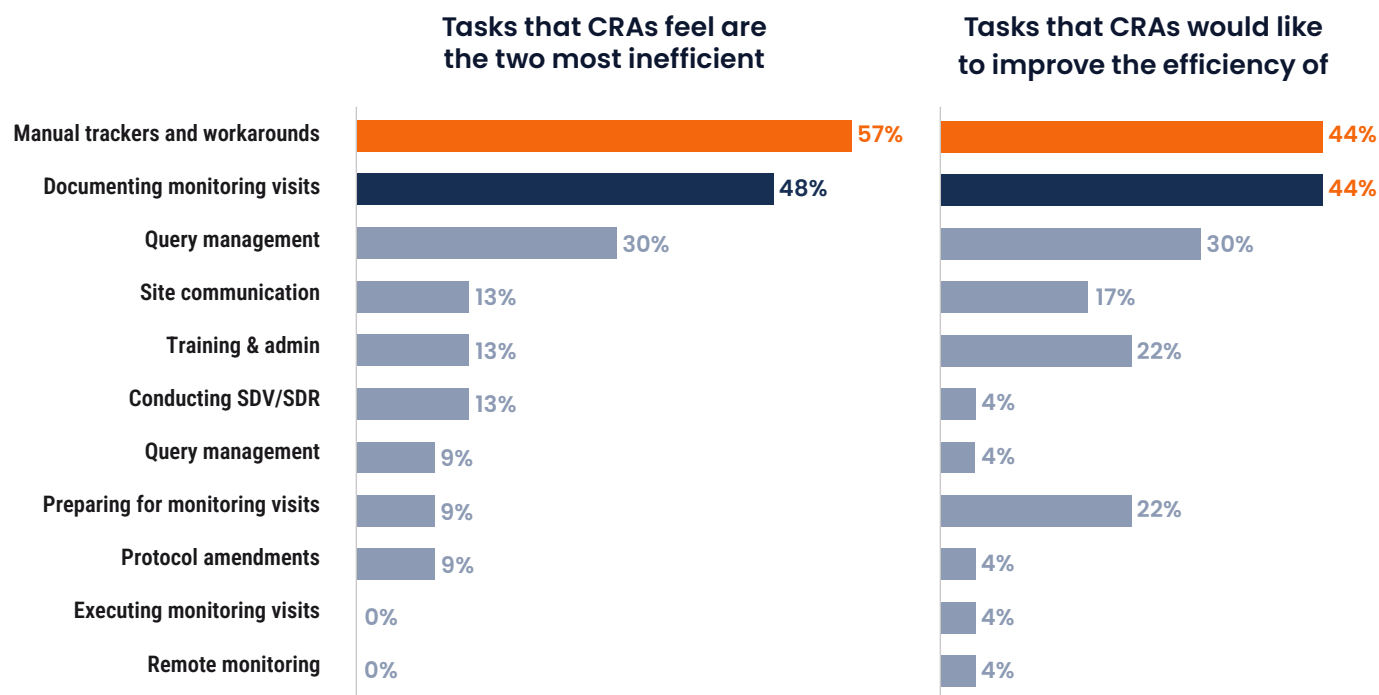


FIGURE 2B



Data captured via online survey from 23 CRAs working on Phase III trials (Due to low sample size, findings may not be statistically significant).

Managing documents in multiple systems and carrying out repetitive steps are major sources of inefficiency. One CRA from a top 20 biopharma expresses surprise that there aren't more automated solutions available: *"We still have to do so much manual tracking, there's no connection between the systems."*

Veeva benchmark analysis from over 30 biopharma companies spanning 10 years shows that a total of 1,286 days per 1,000 monitoring visit reports (MVPs) was wasted. This equates to \$1,028,888, based on an 8 hour work day at a \$100/hour rate. Working in trackers such as Excel, OneNote, and Word accounted for the most time wasted, costing \$326,112 per 1,000 MVPs. If your CRAs complete 20,000 MVPs, for example, the cost would reach \$6,522,240.



I use OneNote for documenting monitoring visits before transcribing into the CTMS. It's not that I like it – I don't have any other tool to keep everything in one place. I have to capture [data] in a non-validated system."

CRA Subject Matter Expert,
Top 20 Biopharma Company

2.3 Cost of queries

Both data managers and CRAs reported that non-EDC queries take far longer (71%) than EDC queries to resolve. Non-EDC queries are those generated from third-party data, such as imaging, eCOA, and central labs. On average, an EDC query takes 9.6 days to resolve, from when it's created to closure, whereas non-EDC queries take 16.4 days. In terms of personal effort on query management, respondents use 34% more active time to resolve a query on third-party data than a query on EDC data. This difference in resolution time is due in part to the delay in non-EDC data reaching the data manager or CRA, and the additional effort required to investigate with the vendor.

It is estimated that a [query costs between \\$28-\\$225 to resolve](#) and that, on average, a Phase III study generates 96,980 queries. Non-EDC queries require 34% more active time, suggesting that their cost ranges from \$32.50-\$300 per query. This suggests that the cost per query is growing to \$32.50-\$300, and that inefficiencies in query management – particularly non-EDC queries – reach millions of dollars in Phase III trials.



Non-EDC queries take longer to resolve in part because data managers aren't receiving the data frequently. When you are querying the data it is months old. The site has to check different systems to find the actual reason or value. We may also keep the query open until the next data transfer, to confirm whether the query was answered correctly or if there is a change on the vendor side."

Senior Clinical Data Manager,
Global CRO

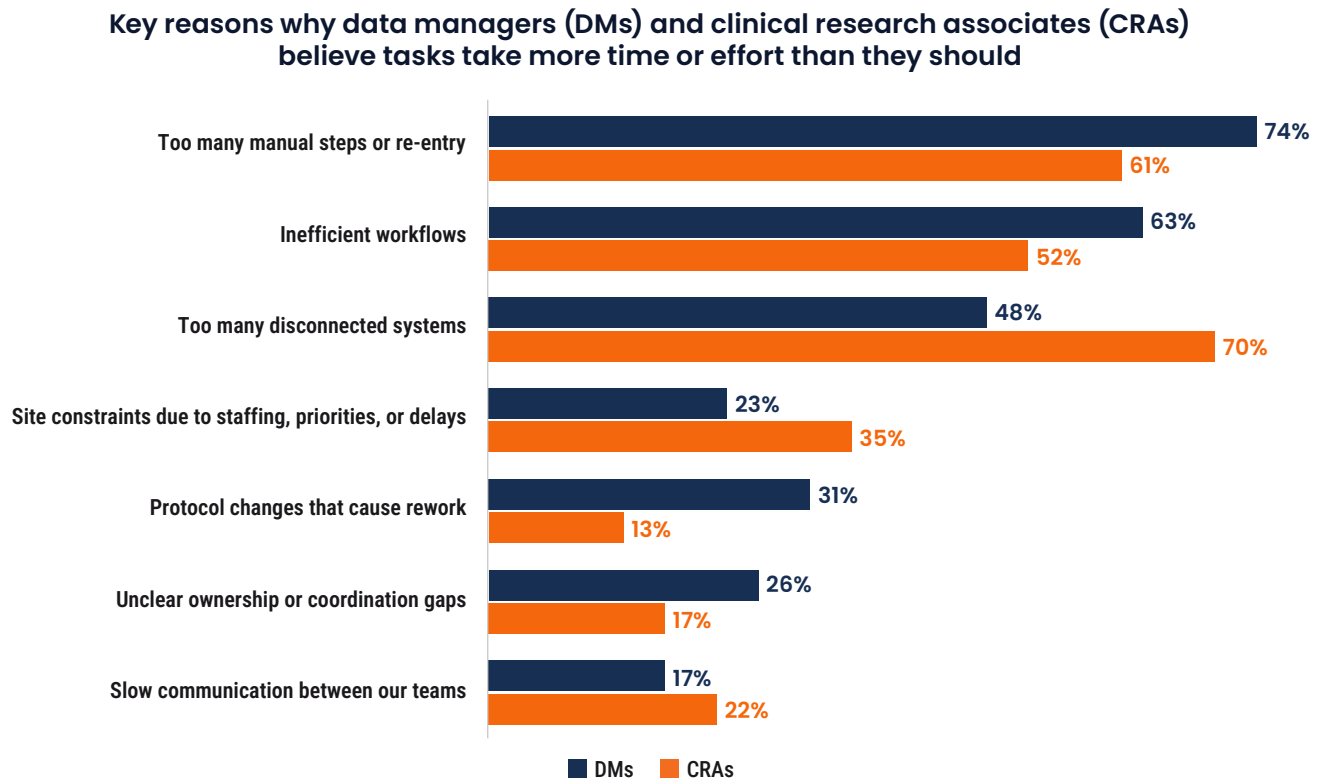
3. Root Causes of Unproductive Effort

3.1 Top drivers of inefficiency for data managers and CRAs

Dealing with disconnected systems is the primary cause of inefficiency for CRAs (71%). Too many manual steps or re-entry (74%) is the main reason why tasks require extra effort and time in data managers' daily work **[Figure 3]**.

One data manager's description of why manual reconciliations are so inefficient highlights the impact of too many manual steps on workflows: "The worst part of data reconciliation is tracking issues from the prior reconciliation to the new data load. Most of us use Excel formulas and manually copy/paste, use Vlookup, carry forward the comment, and then add the latest status. **It is painful.**"

FIGURE 3



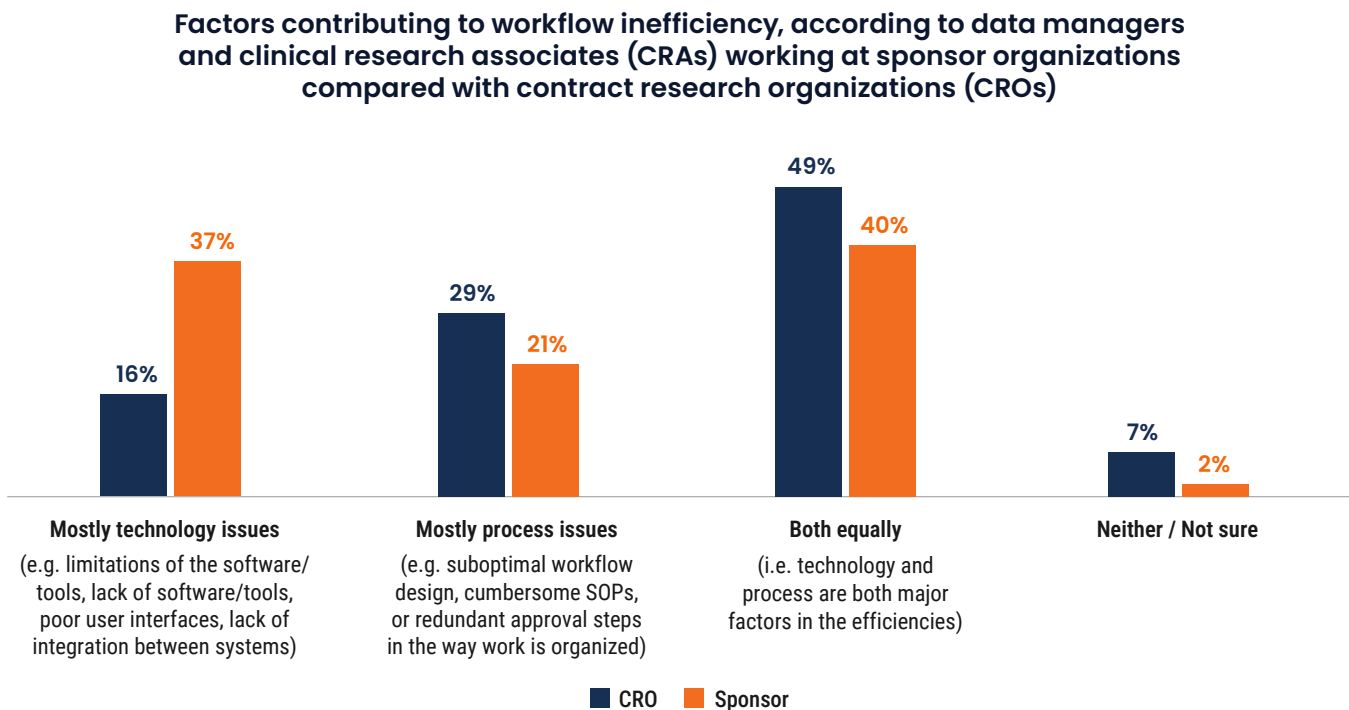
Data captured via online survey from 88 data managers and CRAs (CRO = 45 / Sponsor = 43).

3.2 Tools and processes as sources of inefficiency

When asked whether technology issues or process issues mostly contribute to workflow inefficiencies, or both equally, 48% of respondents said both technology and process issues equally. CRAs were slightly more likely than data managers to point towards tech-specific challenges (30% vs. 25%), as were those working for larger companies (>10,000 employees).

Sponsors are twice as likely to cite technology issues as the primary source of inefficiency than CROs (37% vs. 16%), while CROs more often point to process challenges (29%) or a combination of the two (49%) [Figure 4].

FIGURE 4



Data captured via online survey from 88 data managers and CRAs (CRO = 45 / Sponsor = 43).

4. Implications of Lost Productivity

4.1

Risks and consequences of inaction for data managers and CRAs

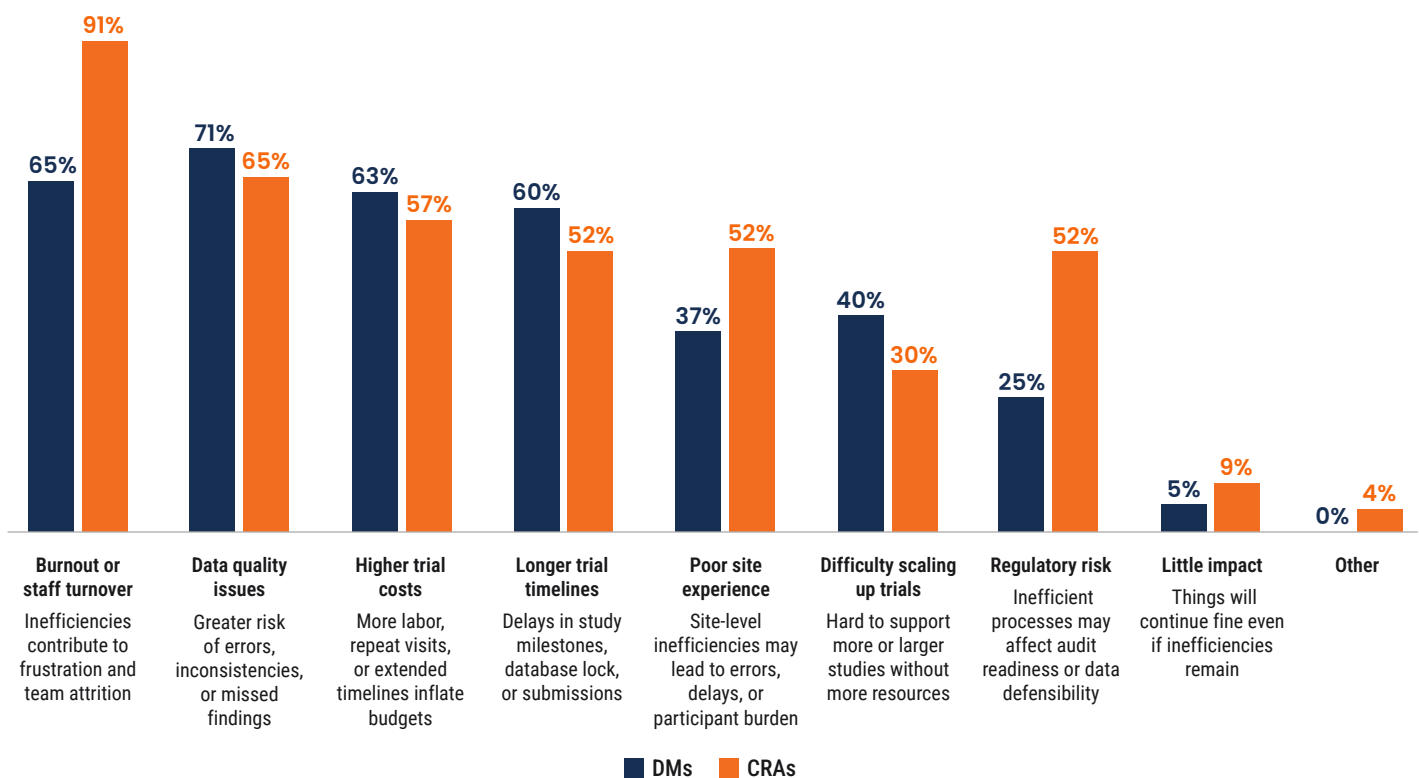
Inefficiencies in Phase III clinical trials have wide reaching impacts, on the trial itself and on those working on the studies. Only 6% of respondents believe that there will be little impact if inefficiencies remain as they are [Figure 5]. For CRAs, the greatest concern is burnout (91%), highlighting an urgent need to act. Clinical data managers also cite burnout as a key risk, but are most concerned with data quality issues (71%).

As risk-based data management approaches become more mainstream, this potentially exacerbates the threat to data quality as teams move away from reviewing 100% of the data. The significant concern over data quality could impede any move towards risk-based reviews, as data managers may be reticent to review anything less than 100% of the data. However, with data volumes growing and burnout risk rising, it is untenable to continue reviewing 100% of the data. Inefficiencies must be addressed for better data quality and data manager upskilling.

As expected, higher trial costs and longer timelines are commonly predicted consequences of inaction. Respondents working at midsize companies are more concerned about higher trial costs (74%) than those at larger companies (52%). These respondents are also more likely to cite regulatory risk as a key concern (42%) compared to larger companies (26%).

FIGURE 5

Anticipated consequences of inefficiencies going unaddressed



Data captured via online survey from 65 data managers and 23 CRAs (Low sample size for CRAs will impact statistical significance of findings).

4.2 Barriers to a more productive future

The biggest challenges to realizing a more efficient future state are protocol complexity (58%), budget/resource constraints (57%), and resistance to change (48%). Data managers were especially likely to cite training gaps (57%), highlighting the need for support as processes and technologies evolve. CRAs were concerned about study complexity and budget/resource constraints equally (65%), with only 22% feeling that training or skills were a barrier.

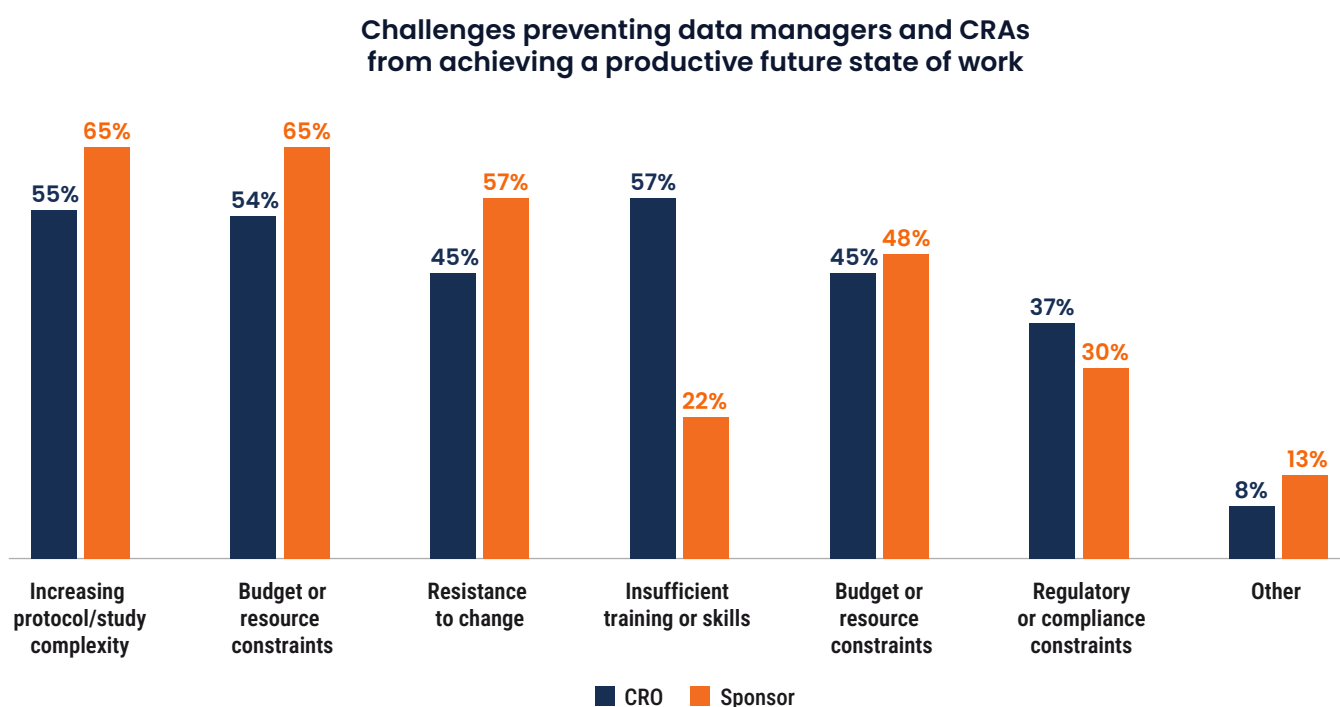
Most respondents (81%) agree that better system connectivity would improve trial execution. Only 57% of CRAs say that their team is actively seeking ways to modernize, compared with 75% of data managers. However, 41% feel that SOPs fail to reflect real-world workflows or underutilize available tools, highlighting a cultural and procedural gap that may slow progress.



Some trials are too complex. If I don't understand something as a user acceptance tester, I have to ask someone else which takes more time than training would in the long run. Lack of training on therapeutic area databases can also result in inadequate testing and incomplete scenario coverage. This can lead to issues that require post-production changes, potentially affecting subject data."

Data manager's response to the question "Why is [manual data reconciliation] so inefficient?"

FIGURE 6



Data captured via online survey from 65 data managers and 23 CRAs. (Low sample size for CRAs will impact statistical significance of findings).

5. The Future Role of Clinical Data Professionals

5.1 Evolution of the clinical data manager role

Data managers have a pragmatic view of how their role will evolve over the next two years. Most expect to use more automation for data cleaning (71%), whereas fewer think AI assistance will support them in decision-making, completing actions, or surfacing data insights (59%). This shows that data managers expect – and prioritize – more automation in their operations before pursuing AI assistance.

Interestingly, fewer expect their roles to shift in a more strategic direction (37%), despite the industry-wide focus on clinical data science, and only 22% think they will be involved earlier in trial design. This suggests that data manager role evolution will be constrained if inefficiencies are not addressed first.

5.2 Evolution of the CRA role

The vast majority (83%) of CRAs expect risk-based monitoring to become the norm over the next two years. One CRA explains their vision for a future-state:



I would love to have a risk-based tool where the monitor can assess whether a site needs more SDV, based on factors like number of PDs, timeliness of data activities, and number of queries. We need good connectivity between systems, because we capture PDs in one system and EDC data in another. A good RBQM tool needs to read across all systems to understand the state of a site."

Compared with data managers (9%), 22% of CRAs think there will be minimal change to their role over the next two years. AI and automation are both seen as likely to impact future roles, as is a greater use of remote monitoring and centralized review (57% each). These shifts point to a future CRA role that's more tech-enabled and focused on critical risk signals, not just site visits.

6. Conclusion and Recommendations

This research highlights a clear message from clinical data managers and CRAs: an unacceptable amount of time is spent on manual tasks and disjointed workflows. These inefficiencies are risking data quality, burnout, and avoidable costs. It marks an executive call-to-action for change, not complacency with past ways of working.

A notable new trend is the burnout risk among data managers – a longstanding challenge for CRAs but a historically less reported concern for data managers. As leaders look towards increasingly innovative data management approaches, this research reminds us that the people doing the work are not enabled by modern technology or efficient processes. We recommend a practical approach:



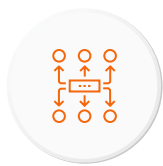
1. Prioritize system integration

Invest in solutions that connect clinical data and operations systems. This will reduce manual data entry, and provide a more unified view of trial progress.



2. Embrace automation

Centralize data review and reconciliation efforts in an automated tool that directly addresses the most time-consuming and frustrating manual tasks, within a system connected with clinical operations.



3. Streamline workflows

Review and update SOPs to ensure they reflect real-world workflows and fully leverage available technologies. This may involve revisiting processes that are overly complex or rely on outdated manual steps.



4. Invest in ongoing training and change management

Provide adequate training and ongoing support to data managers and CRAs for long-term adoption of new tools and processes. This will help overcome resistance to change and ensure that professionals can effectively use new systems.



5. Foster a culture of efficiency

Encourage a mindset that values efficiency and continuous improvement. This means actively seeking feedback from clinical professionals and empowering them with the tools and processes they need to succeed.

Discover how others are optimizing traditional processes in this **video**.



7. Appendix

7.1 Methodology

RESEARCH DESIGN

- | Mixed-methods approach combining quantitative surveys and qualitative interviews
- | Rationale for chosen methodology: to quantify broad trends and capture in-depth experiences

QUANTITATIVE SURVEY

Participants

- | **Target audience:** Data managers and CRAs
- | **Sample size:** 88 total participants (65 data managers, 23 CRAs)
Important note: The small CRA sample size should be interpreted with caution. We have called out instances where considerable differences exist and sample sizes are robust enough, but these should still be interpreted with caution. Overall, findings are very consistent across segments.

| Geographic spread:

	Total	DM	CRA
North America	35%	35%	35%
Europe and UK	21%	12%	44%
Asia Pacific	32%	40%	9%
Other	12%	12%	13%

- | **Survey instrument:** Online questionnaire
- | **Data collection:** The survey was open for a total of 70 days, between May 2 and July 7 2025

QUALITATIVE INTERVIEWS

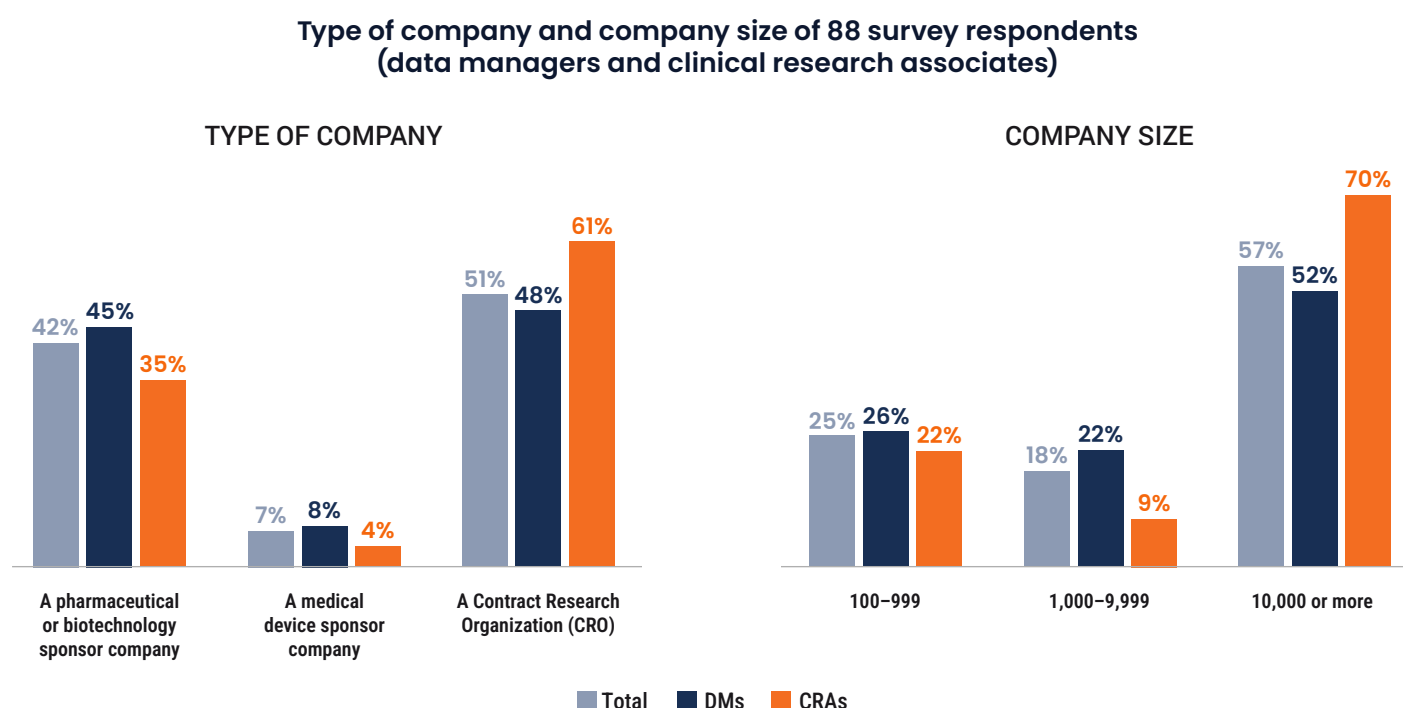
- | **Participants:** Five survey participants took part in the qualitative interviews, with the following profiles:
 - Three data managers from CRO companies
 - One data manager from a sponsor company
 - One CRA from a sponsor company
- | **Interview process:** Semi-structured interviews, approximately 60 minutes long
- | **Data collection:** Interviews were carried out between June 25 – July 17 2025

7.2 Survey demographics

Respondents bring deep experience to their roles, with over half having more than 10 years of experience and strong involvement in Phase III trials. Data managers are especially tenured, while CRAs show higher trial volume – nearly half have supported 10 or more Phase III trials in the past five years.

Respondents represent a mix of sponsor companies and CROs, with CROs being the most common employer, especially among CRAs. Most work at large organizations: over half are based at companies with 10,000 or more employees, suggesting well-established operational environments.

FIGURE 7



When asked which clinical trial software providers' solutions they have used in their work in the last year, most typically use a varied set of solutions: Rave EDC, Veeva eTMF, Veeva EDC and Veeva CTMS were the most used tools reported by respondents. Rave EDC was used by 75% overall, with CRAs reporting usage of Veeva eTMF (68%), Veeva CTMS (78%), and Oracle Siebel CTMS (30%).

7.3 Full quantitative survey results

FIGURE 8

Data managers' response to the question: "For [tasks ranked as the two most inefficient], how is that task primarily carried out in your current trials?"

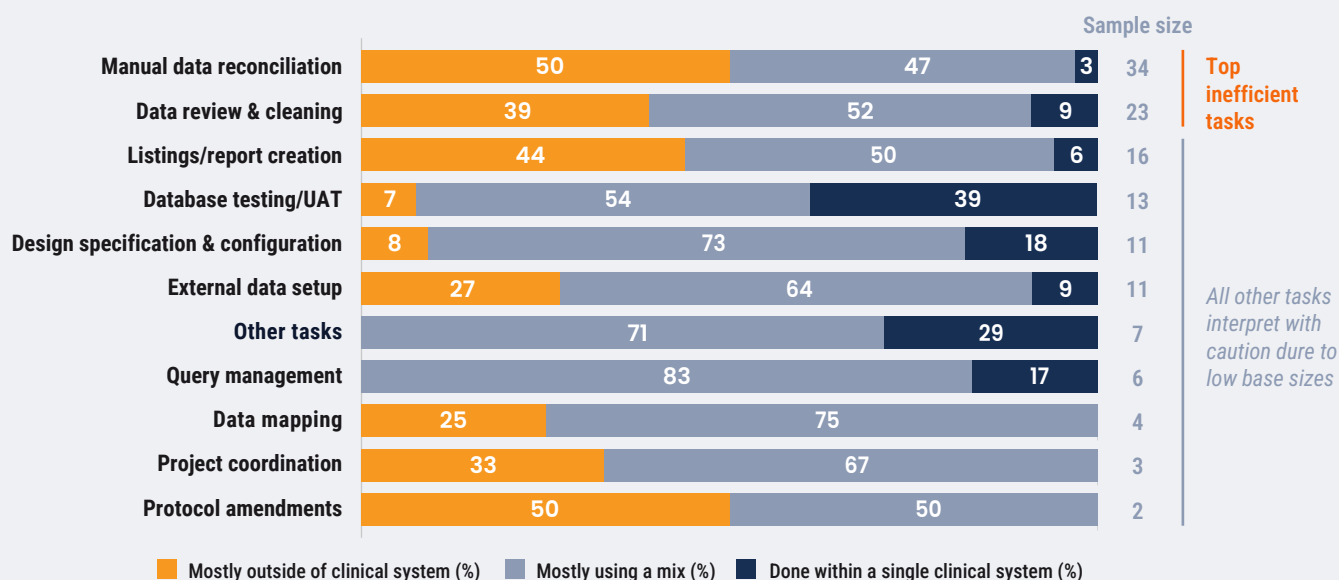


FIGURE 9

CRA's response to the question: "For [tasks ranked as the two most inefficient], how is that task primarily carried out in your current trials?"

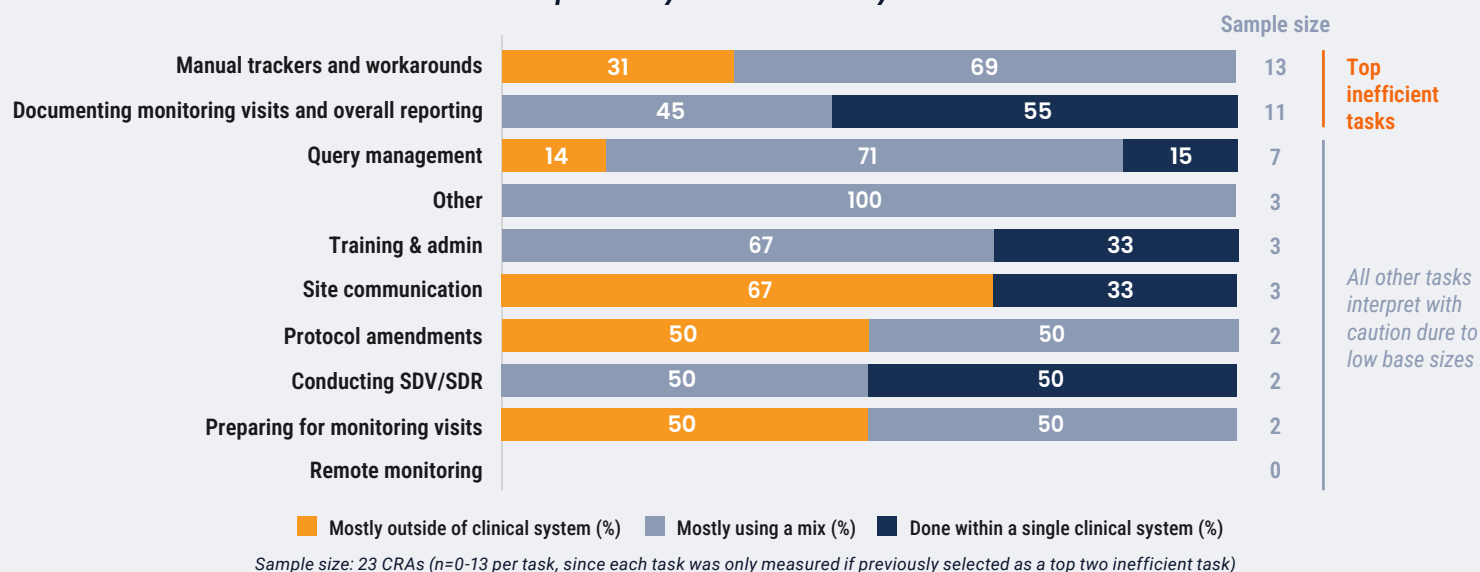
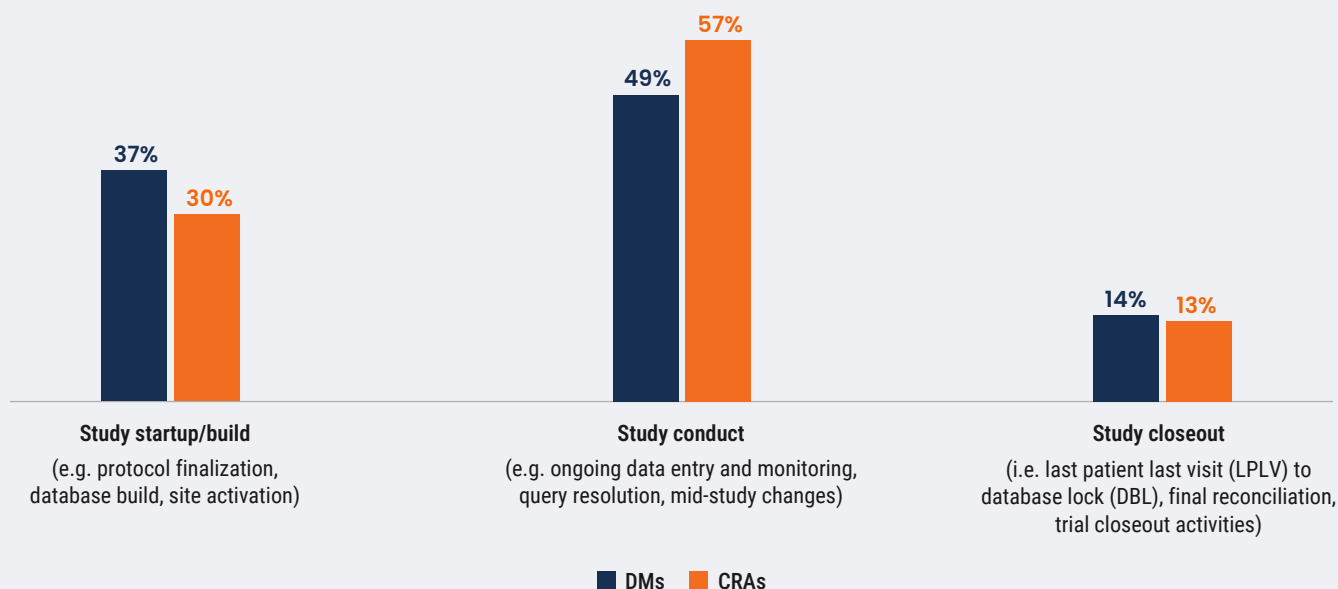


FIGURE 10

Data manager and CRA responses to the question: “In your experience at your current company, during which phase of a typical Phase III study do you see the most inefficiency?”



Sample sizes: Total 88: 65 data managers; 23 CRAs (Low sample size for CRAs will impact statistical significance of findings)

FIGURE 11

Data manager and CRA responses to the questions:

- “On average, how much total elapsed time (in days) does it take to resolve a single EDC and non-EDC data query – from when it’s created to when it’s closed out? This would include both your active time working on it AND any time spent waiting for responses from sires, vendors, or internal teams.”
- “On average, how much of your personal time (in minutes) do you spend actively managing and resolving a single EDC and non-EDC data query? This would include time spent creating, tracking, sending, following up on, and closing out the query. It would NOT include time spent waiting for responses from others.”

TOTAL

Average time to resolve a query
(elapsed days)

9.6 days **16.4 days**
EDC Non-EDC

Non-EDC queries take **71%** more time

DMs

Average time to resolve a query
(elapsed days)

9.7 days **16.7 days**
EDC Non-EDC

CRAs

Average time to resolve a query
(elapsed days)

9.3 days **15.7 days**
EDC Non-EDC

Personal effort to resolve a query
(active minutes)

15.2 min. **21.1 min.**
EDC Non-EDC

Non-EDC queries take **34%** more time

Personal effort to resolve a query
(active minutes)

15.1 min. **21.2 min.**
EDC Non-EDC

Personal effort to resolve a query
(active minutes)

15.5 min. **20.7 min.**
EDC Non-EDC

Sample sizes: Total 88: 65 data managers; 23 CRAs (Low sample size for CRAs will impact statistical significance of findings)

FIGURE 12

Data managers' response to the question: "For each of the following tasks you perform during Phase III trials, how would you describe the time it currently takes compared to how long it should take in an ideal, efficient workflow?"

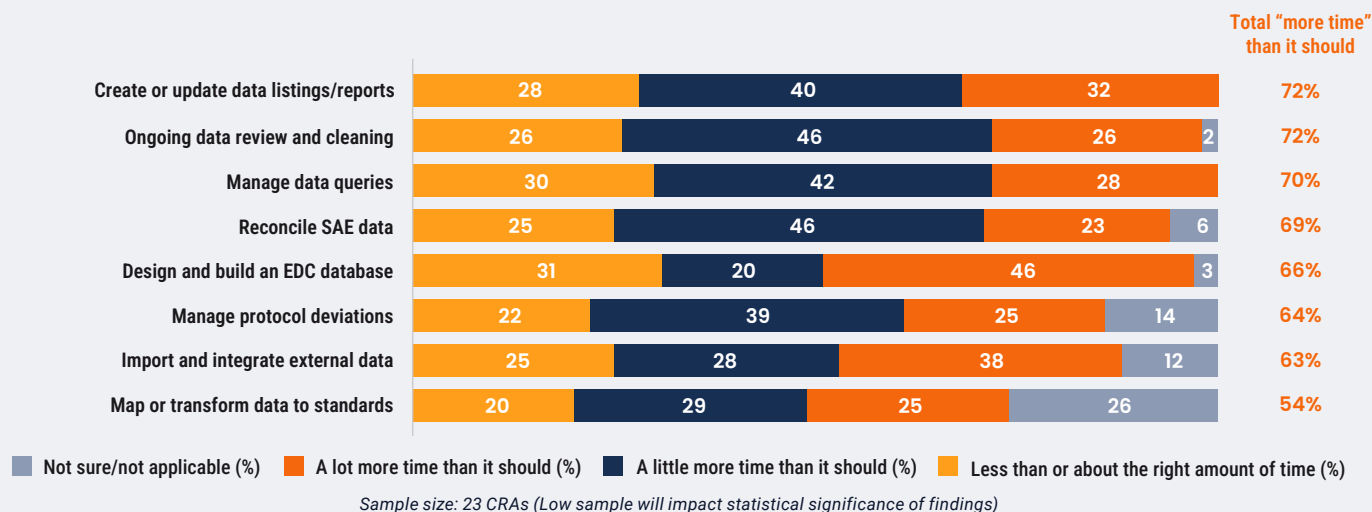


FIGURE 13

CRAs' response to the question: "For each of the following tasks you perform during Phase III trials, how would you describe the time it currently takes compared to how long it should take in an ideal, efficient workflow?"

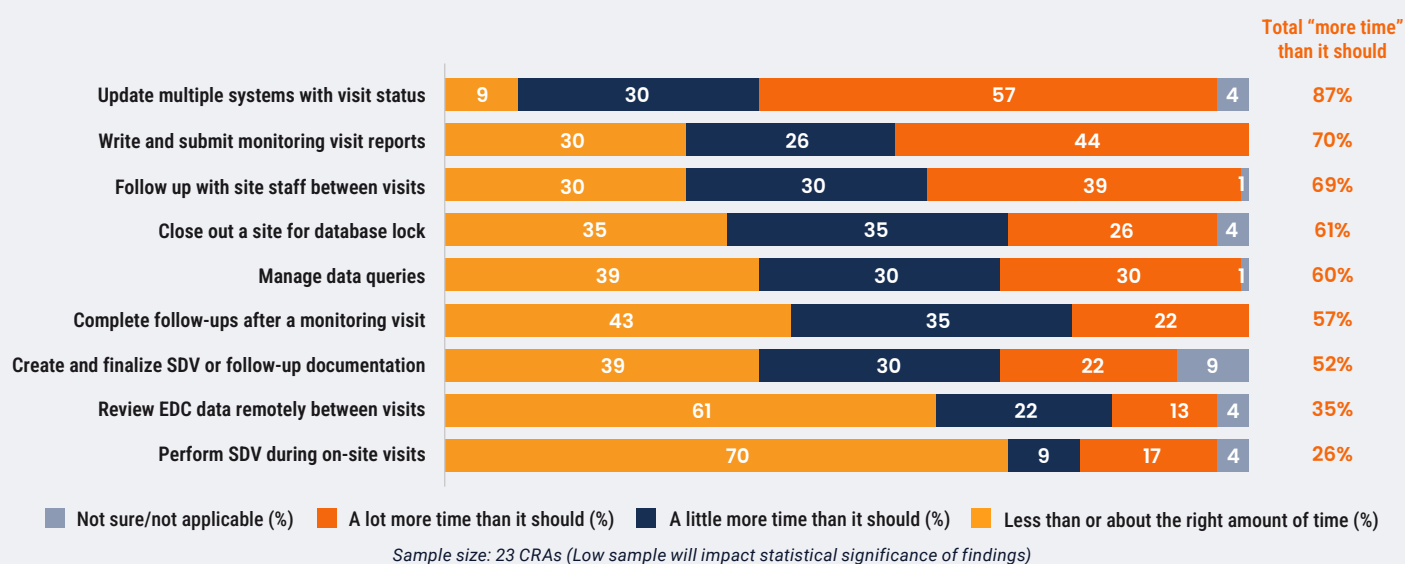
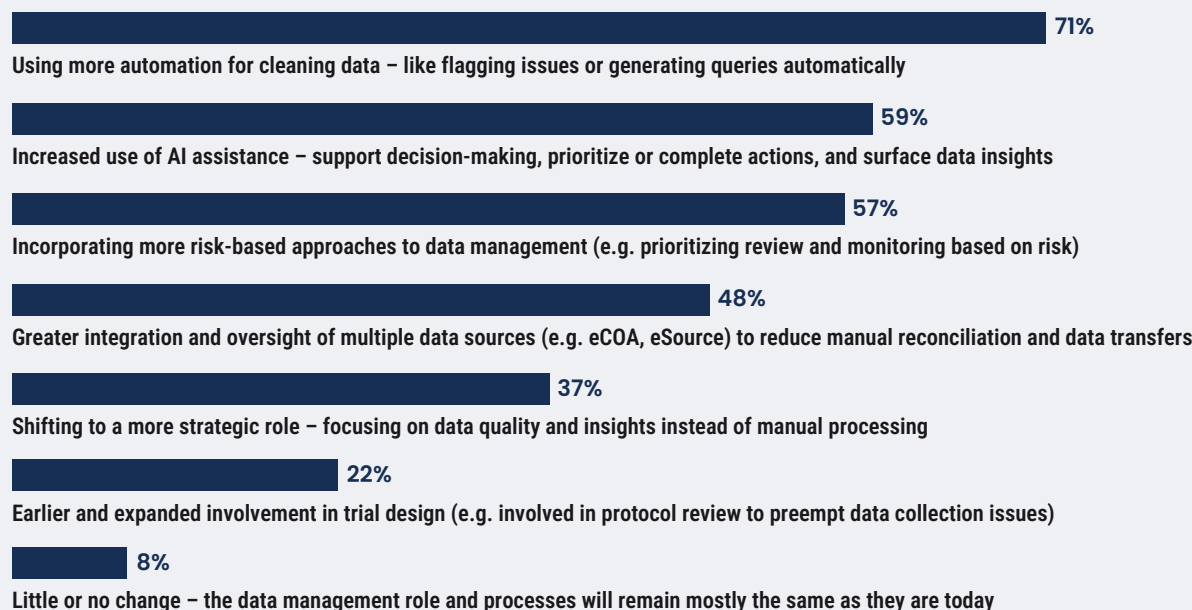


FIGURE 14

Data managers' response to the question: "How do you most see your role and its processes evolving over the next two years?"



Sample size: 65 data managers

FIGURE 15

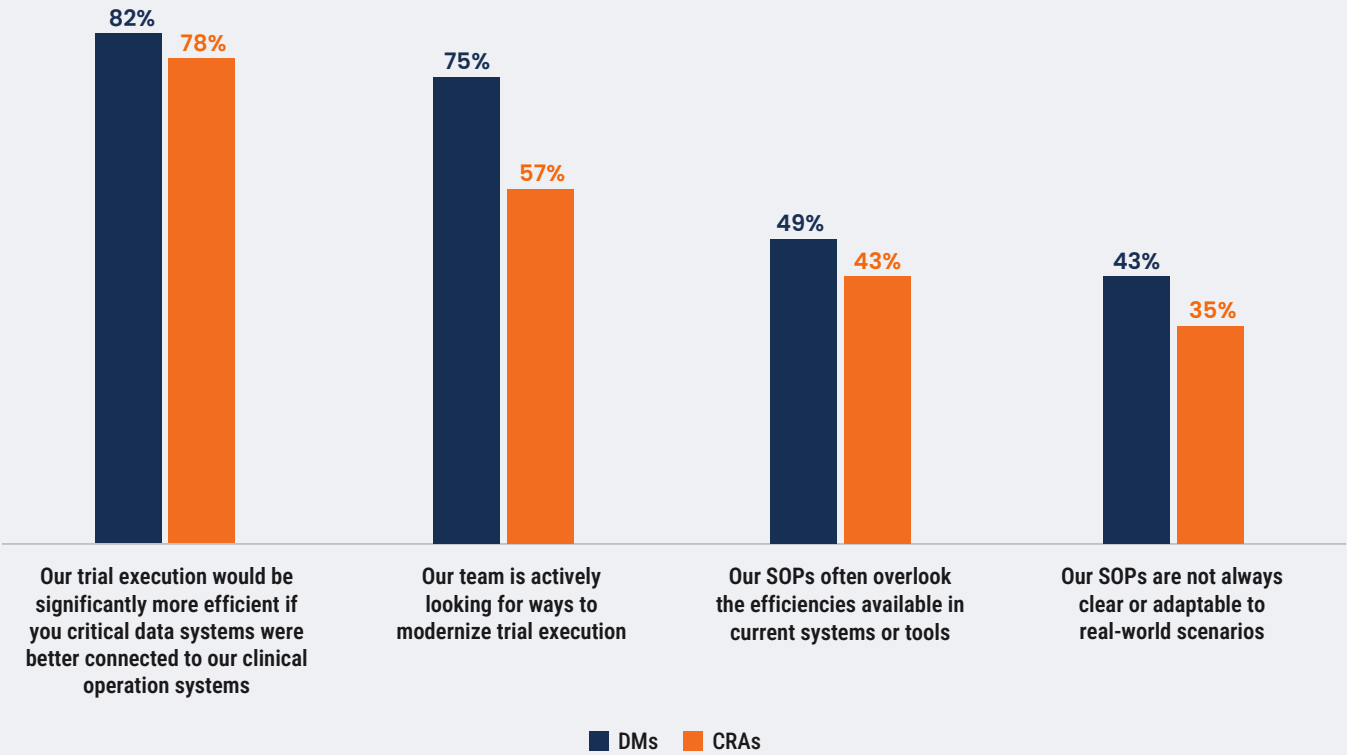
CRA's response to the question: "How do you most see your role and its processes evolving over the next two years?"



Sample size: 23 CRAs (Low sample will impact statistical significance of findings)

FIGURE 16

Data manager and CRA responses to the question: “How much do you agree or disagree with the following statements? (Strongly disagree, Disagree, Neither disagree or agree, Agree, Strongly agree).”
Chart shows the top two levels of agreement.



Sample sizes: Total 88: 65 data managers; 23 CRAs (Low sample size for CRAs will impact statistical significance of findings)

8. Acknowledgements

Veeva would like to thank its partners at SCDM for distributing the survey and sharing this report with its members.



ABOUT US

Veeva is the global leader in cloud software for the life sciences industry. Committed to innovation, product excellence, and customer success, Veeva serves more than 1,100 customers, ranging from the world's largest biopharma companies to emerging biotechs. As a Public Benefit Corporation, Veeva is committed to balancing the interests of all stakeholders, including customers, employees, shareholders, and the industries it serves.

For more information, visit www.veeva.com