Veeva

2025 Clinical Data Trend Report

Table of Contents

Outlook for 2025:
New forces driving pragmatic innovation

In clinical data, there is often a lag between a new technology becoming available and its widespread adoption. For instance, it took about a decade before most of our industry regularly used electronic data capture (EDC) during clinical trials instead of paper.

Instead of signifying caution about embracing innovation, these delays show that it just takes time to define a credible value proposition that overcomes risk aversion in our highly regulated industry. However, today we believe there are six key trends to follow closely, each of which already has real use cases. Along with four emerging trends, we anticipate they will reshape clinical data in years to come.

A unifying theme across these 10 trends is the notion of "simplify and standardize". This applies focus to the increased complexity in the clinical landscape and drives us toward pragmatic innovation.

Some of these trends are interconnected. For example, the shifts to clinical data science and endpoint-driven design are foundational to the rise of risk-based approaches; otherwise, data teams wouldn't know which data points to focus on. Others, such as the pivot to smart automation after AI hype and the emerging perspective on metadata repository-driven (MDR) builds, reflect where companies want to focus their energies. Many used AI pilots to test, fail, and learn and now believe that a mix of rule-driven and AI-based automation will deliver the most significant cost and efficiency improvements.

As the debate on decentralized clinical trials (DCTs) moves on, there is also a renewed emphasis on patient optionality and centering innovation around better site experiences. These areas are critical if the industry is to respond to recent FDA guidance encouraging 'pragmatic trials' in certain scenarios. By incorporating design elements that more closely reflect routine clinical practice, the hope is that more patients (including those from diverse populations) will want to enroll and contribute to clinical research.

We hope this report helps your teams ideate, plan, and prioritize their clinical data initiatives so that we can deliver better trials for all.

Drew Garty
Chief Technology Officer
Veeva Clinical Data

Six trends reshaping clinical data

1. The rise of risk-based everything

Regulators have long encouraged risk-based approaches to quality management (RBQM) and are now applying the same principles to data management and monitoring. ICH guideline E8(R1) asks sponsors to consider critical-to-quality factors in clinical research and manage "risks to those factors using a risk-proportionate approach".[1]

Given ever-expanding data volumes, it is not sustainable for biopharma companies to scale data management linearly using traditional methodologies. With regulators now supporting risk-based approaches, teams are shifting focus from traditional data collection and management activities to dynamic, analytical tasks. Rather than reviewing

everything, they concentrate on the most important data points. Instead of 'marshaling' data, their aim is to generate valuable insights (see 'Endpoint-driven design enables risk-based approaches').

Although interest in risk-based approaches is high, most are yet to make the leap of faith and move away from the security blanket of comprehensive review models. Still, clinical leaders believe risk-based approaches will add value quickly, so are upskilling their data managers to transition to clinical data science while putting in place more advanced technology **[Figure 1]**.

FIGURE 1
RBQM is considered most likely to succeed and deliver value in the near term
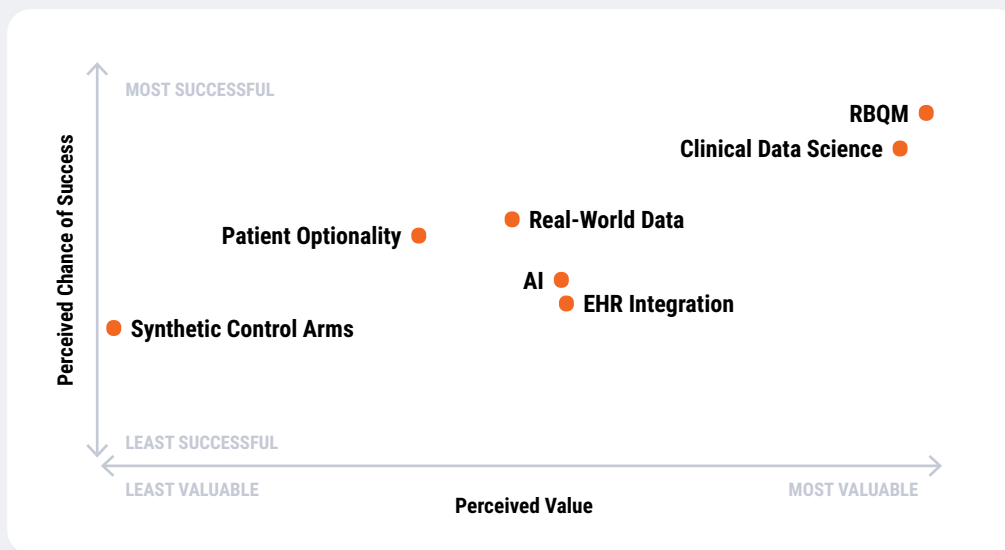
Perceived Value vs. Chance of Success for Clinical Initiatives

Source: Veeva Systems. Represents responses from 156 clinical data leaders who attended roundtables in New York, London, Basel, Copenhagen, and the Bay Area in 2024. Answer to question: "Which initiatives have the highest probability of success AND the highest value in the next 2 years?"

[1] European Medicines Agency, 'ICH guideline E8(R1) on general considerations for clinical studies', 2022

REAL USE CASES

At a global biopharma, combining risk-based checks with technology is essential for focusing on critical data. As clinical research associates (CRAs) can now see source data verification (SDV) requirements, they no longer need to download reports or apply macros in spreadsheets. "Eliminating one 20-minute task per visit across 130,000 visits avoids 43,000 hours of work. CRAs can focus on what matters," explains a senior director within the global biopharma's clinical sciences and study management group.

Similarly, straightforward improvements to system functionality — such as not requiring users to enter future visit dates — will avoid an estimated 54,000 queries a year **[Figure 2]**.

EMERGING USE CASES

Some sponsors are experimenting with historical trend data for proactive issue management. After defining thresholds and sharing data across departments, they assess how a trend changes over time, communicate their findings, and document issue remediation.

To do this, you need to run fit-for-purpose methods from your previous trials. Then, align as a cross-functional team on critical risks and gather input early from your study team **[Figure 3]**. Some of the risks you identify can be tolerated in the trial as long as you have mitigation plans and procedures in place. Once the trial starts, empower a centralized team to review and monitor data: they will be able to identify signals and data anomalies as they come in, and surface information to the right people.

Risk-based everything introduces value-creation opportunities to your trial:

1. **Higher data quality** (due to proactive issue detection) leads to faster approvals
2. **Greater resource efficiency** (through centralized data reviews) reduces trial costs
3. **Shorter study timelines** (from reduced time to database lock) speed time to market
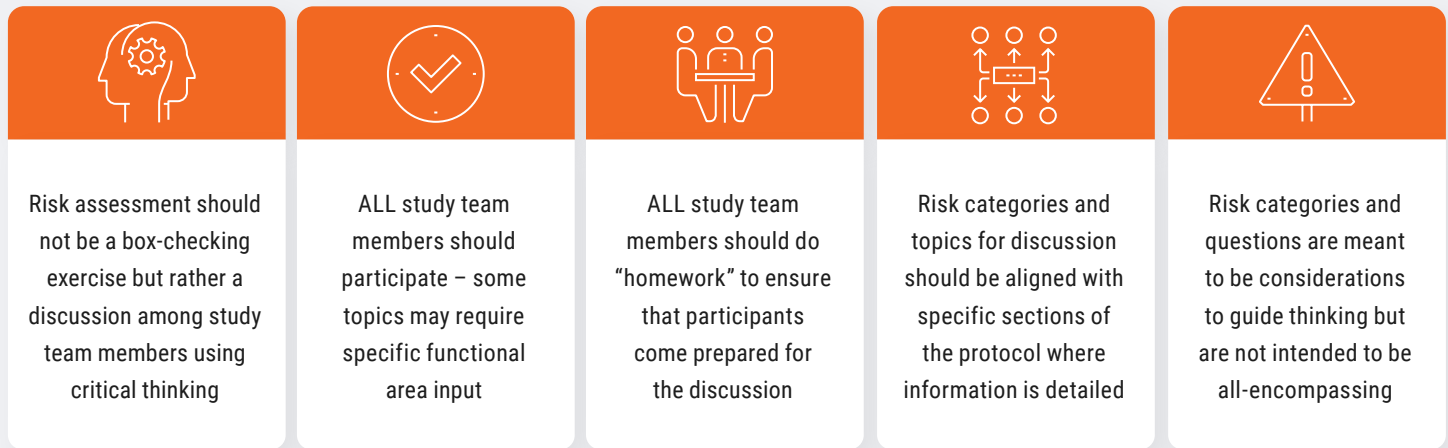
> "
>
> *It's critical to plan as a cross-functional team to get the right risk assessment in every trial. Push critical thinking into the study team and ensure all stakeholders are aligned on the terminology. The right balance lies between a rigid questionnaire and a freeflow discussion."*
>
> **Dan Beaudry**
> Senior Vice President, Product Management, CluePoints

Structuring a productive risk-mitigating discussion

Risk assessment should not be a box-checking exercise but rather a discussion among study team members using critical thinking	ALL study team members should participate – some topics may require specific functional area input	ALL study team members should do "homework" to ensure that participants come prepared for the discussion	Risk categories and topics for discussion should be aligned with specific sections of the protocol where information is detailed	Risk categories and questions are meant to be considerations to guide thinking but are not intended to be all-encompassing

Source: CluePoints

2. Clinical data management evolves into clinical data science

Three years ago, it was predicted that **clinical data science would one day emerge from clinical data management.** One year ago, **theory started to become reality:** asked to rank which initiatives have the highest potential value and likelihood of success over the next two years, clinical leaders ranked data science second **[Figure 1]**.

Because more activities can now be automated, data managers can shift their focus from operational tasks (such as data collection and cleaning) to strategic contributions (e.g., generating insights and predicting outcomes). But the shift from managing data to applying data scientifically introduces new challenges. Clean, harmonized data is increasingly a 'product' for downstream groups and other consumers, with data managers marshals of this data.

Breaking down the barriers between data management and other functions is a prerequisite for clinical data science. When data managers partner effectively with clinical operations and safety, for example, the whole organization benefits from streamlined end-to-end data flows and better decision-making. As more sources of patient data come into play, data managers will need to "stop babysitting data that doesn't matter".[2]

[2] New York Panelist, Clinical Data Innovation Forum

Clinical data leaders note that **the shift from data management to clinical data science is underway** and requires new KPIs. Their focus areas include optimizing patient data flows, using AI/ML and advanced analytics, integrating data quality and review, and automating analysis. To achieve all of this, data managers need to **develop new skill sets.** As one leader notes, "We need to move away from merely checking boxes toward interpreting data".[3]

The mission of an exceptional data scientist remains the same as before: extract the most value from ever-expanding volumes of data. For this to be feasible, they must be involved in early trial design and protocol development. The rise of 'risk-based everything' and endpoint-driven design (covered in this report) will enable data scientists to analyze the right data.

Watch Pfizer's Demetris Zambas discuss whether data science requires data perfection

3. Focus shifts from AI hype to smart automation

'Smart automation' aims to leverage the best automation approach — whether AI, rule-based, or other — to optimize efficiency and manage risk for each specific use case. Simply put, it puts value creation ahead of a shiny label.

Many biopharma companies are trying to formalize their AI initiatives. However, in discussions with industry leaders, AI is ranked as less likely to succeed or deliver value in the medium term than other initiatives **[Figure 1]**. Because AI is a 'black box' solution, it's difficult to predict how it will behave, which tilts the risk calculation unfavorably. Justifying its outputs can be problematic so, depending on the task, companies will need 'humans in the loop' to review (and decide on) AI-generated work.

In contrast, human oversight isn't required for rule-based automation. Companies that realize this are investing pragmatically in automation, introducing capabilities that add real value today while establishing the infrastructure for future AI use cases.

Generating real value today:

1. **Standardized data acquisition.** Applied across sources, this automates data import and mapping to significantly reduce average study times
2. **Rule-driven automation.** Speeding up data cleaning, transformation, and reporting enhances data trust and reduces manual work

Enabling emerging and future AI use cases:

1. **High-velocity APIs**
2. **Feedback loops**, where system reactions are fed to AI engines, will help them learn

[3] New York Panelist, Clinical Data Innovation Forum

EMERGING USE CASES

Of the AI-augmented solutions that are more likely to succeed in delivering value in the near term, medical coding stands out as a clear leader. AI augmentation fits neatly into a slightly modified medical coding workflow **[Figure 4]**. Traditional rule-based automation is already in place today and accounts for most of the automation of medical coding. For records that do not get automatically coded, AI can be applied to either offer a medical coder a suggestion or to automatically code and have the medical coder review the selected term.

Remember: your AI-based solution needs high-context data to solve high-context problems. Imagine you want to use AI to identify trial subjects who have a worsening condition but no associated medical history. This is a high-context problem. If the AI-based solution's only context is its training on 'low-context' past data, it will not be able to solve this problem accurately and reliably. The AI solution would still need human review and feedback to judge whether or not to create a query.

> *We did a small AI initiative to see if we can generate meaningful test data for setting up and validating our systems. It turns out we can. An algorithm that we developed looks at past studies we set up, learns from the actual data collected, and uses it to generate something we can use for new studies."*
>
> **Ibrahim Kamstrup-Akkaoui**
> Vice President, Clinical Data Operations, Novo Nordisk
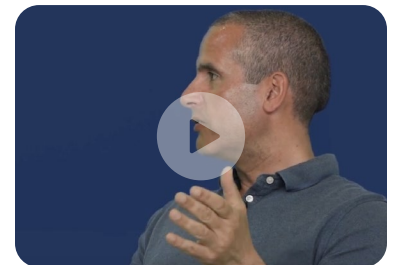
Watch the full episode

FIGURE 4
AI-augmented medical coding
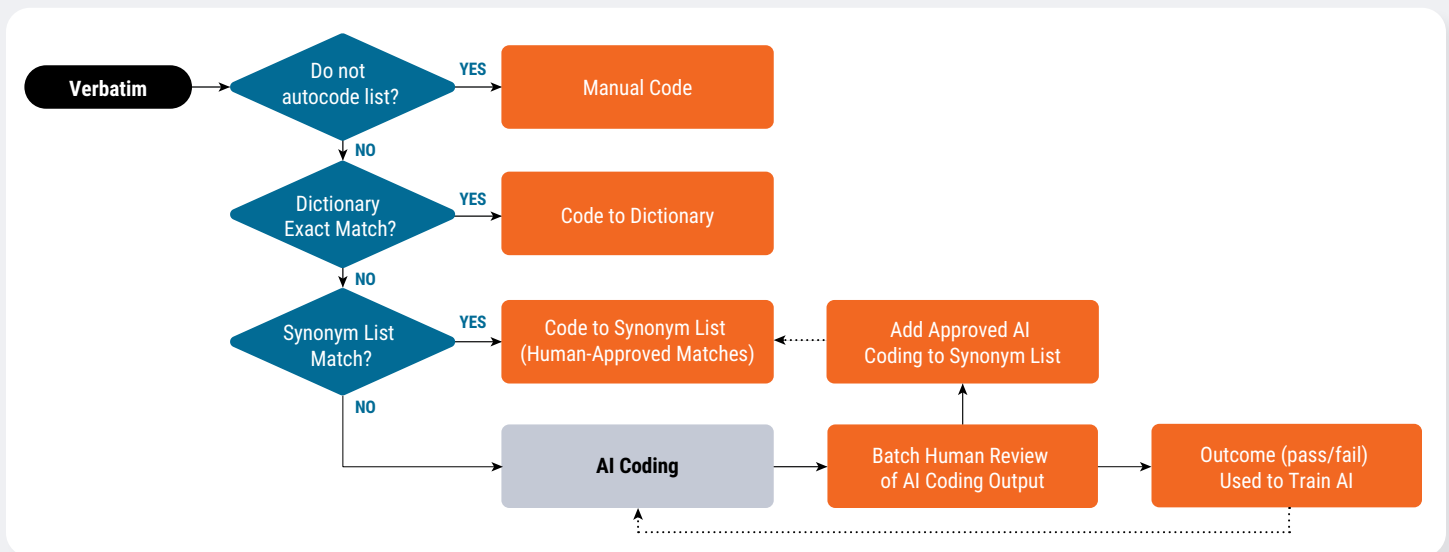
In comparison, a rule-based approach can automate the same processes without human review by using the following ruleset:

1. Identify every patient record with a worsening condition report

2. If any of these records do not have an associated ongoing medical history, add it to a list and create a query

Today, biopharma companies like **GSK are already using rule-based automation for data cleaning to accelerate the time to database loc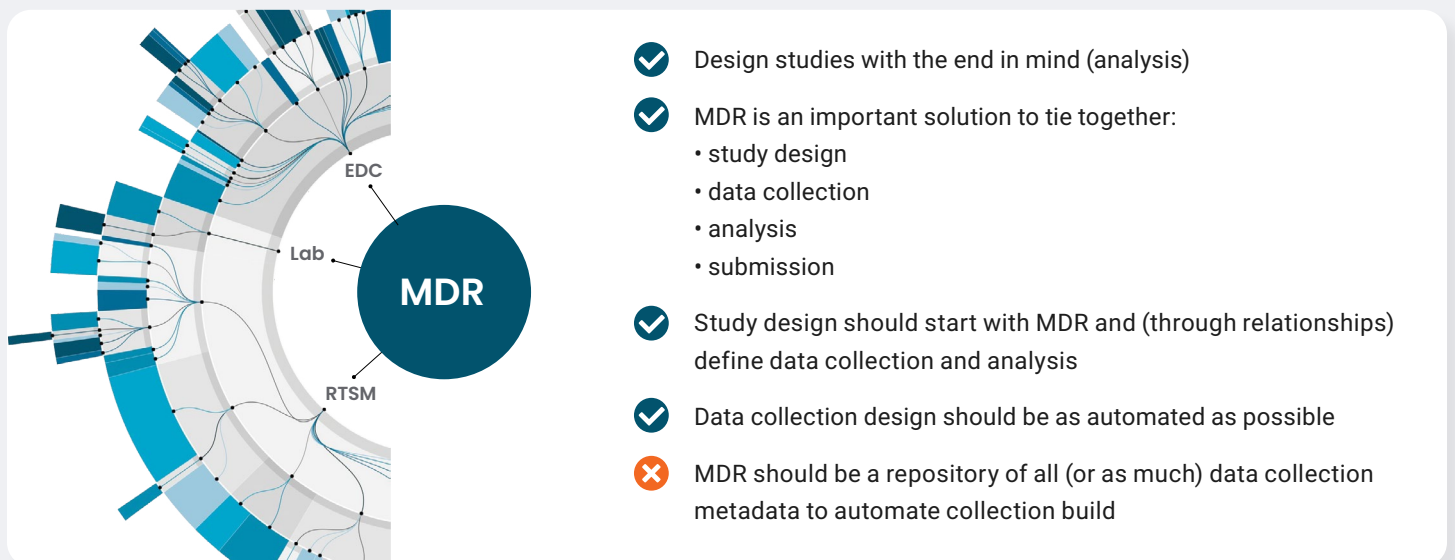k.** In the medium term, we believe rule-based automation will drive the most significant cost and efficiency improvements. Later, companies envision GenAI becoming a co-pilot during clinical studies: perhaps providing prompted insights or suggestions, detecting fraud, or predicting compliance adherence.

A clean data foundation, built with smart automation, will enhance the quality (and expedite the delivery) of the data required to power AI use cases further down the line.

4. The resurgence of MDR and data standards

Metadata repository-driven (MDR) builds are not new, but the approach is becoming more pragmatic by focusing efforts on specific business problems **[Figure 5]**.

Focusing MDR on what matters

- Design studies with the end in mind (analysis)
- MDR is an important solution to tie together:
  - study design
  - data collection
  - analysis
  - submission
- Study design should start with MDR and (through relationships) define data collection and analysis
- Data collection design should be as automated as possible
- MDR should be a repository of all (or as much) data collection metadata to automate collection build

Source: Veeva Systems

MDR solutions help tie together study design, data collection, analysis, and submission. Once EDC became the principal system in clinical data collection, the prevailing view was there should be one repository for all (or almost all) data collection metadata to automate study builds. In reality, it proved challenging for companies to scale metadata management, particularly because many still rely on spreadsheets.

REAL USE CASES

Instead of a repository for all metadata, a more effective emerging strategy is to focus MDR on what matters: the study design metadata that are common, shared, and critical to data management and statistics. For instance, when looking for common study design metadata between data collection and data analysis, there might be as few as 25 properties (out of more than 1,000) of EDC metadata that affect downstream programming and analysis. The study design would start with MDR, and agree standardized data definitions at the data collection stage. Data management and stats could then work in parallel on delivering to the same definition.

Moving from 'big', all-encompassing MDR to simplified standards will accelerate the path from study build to database lock. Traditional MDR and spreadsheets slow down companies whereas a pragmatic approach will mean they can deliver value faster. For example, **Faro Health,** a generative AI company, can now automate the creation of an EDC study build in just seven API calls.

5. Patient optionality is now the real goal of DCT

It is estimated that only three percent of U.S. physicians and patients take part in clinical trials that lead to new therapies. One consequence of low participation is that nearly 80% of trials fail to meet enrollment timelines, causing costly delays.[4]

The **Food and Drug Administration's (FDA) definition of decentralized clinical trials** is commonly used and high level: "A clinical trial that includes decentralized elements where trial-related activities occur at locations other than traditional clinical trial sites." In recent years, the DCT debate led to hyper-focus on where research takes place without due consideration to *how* **it impacts the overall trial experience** (e.g., on patients, sites, data managers, regulators, etc.)

Instead of focusing on where, clinical data leaders are prioritizing patient optionality: DCT technology is now a standard way of operating and supporting the patient experience **[Figure 6]**. Putting patients in control of how they participate in a trial — whether at home, at a site, in a clinic, or another care setting — will result in more timely and cost-effective clinical research.

While introducing more technology to trials has some benefits, sponsors are thinking holistically about the trial experience: otherwise, patients could easily become overwhelmed by the number of devices they are asked to use and wear. 'Bring your own device' **(BYOD) policies are a good way of making clinical research more convenient** while maintaining data quality and security.

[4] FDA data, quoted in Fierce Healthcare

Perceived value of technologies that expand patient optionality
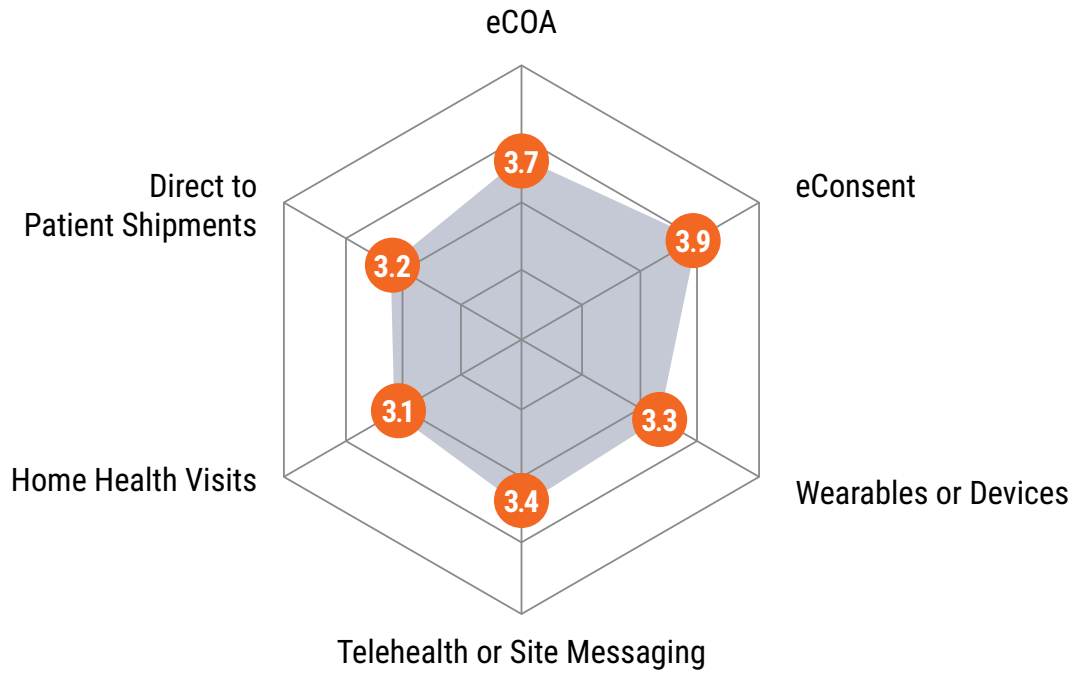
Source: Veeva Systems. Represents responses from 55 clinical data leaders who attended a roundtable in Boston in 2024. Answer to question: "On a scale of 1-5, what is the value of these technologies?"

REAL USE CASES

Some clinical data leaders are alleviating the patient burden at the protocol design stage, by asking study participants for less data. Others are considering whether there are tangible patient benefits before introducing new applications (e.g., eConsent) or using surveys to understand the patient experience and pinpoint improvements.

6. Innovation focuses on better site experiences, at scale

In recent years, trials have evolved from every patient visit occurring at a site to greater visit method optionality within protocols **[Figure 7]**.

Modern trial methods require sponsors to aggregate digitized and non-digitized information from lab feeds, diagnostics, medical devices, and electronic health records (EHR). While **clinical workbenches centralize data cleaning, aggregation, and reconciliation** for sponsors, sites currently lack access to a centralized database for all their relevant clinical data sources.

Instead of a linear site-to-sponsor data flow, sites now navigate hybrid data flows and collect and verify diverse data sources. For instance, if a patient chooses to attend a site for their first visit, and then goes to a clinic (or stays at home) for their second, their data will flow through electronic data capture (EDC) for their first visit, and another system for their second.

At its best, innovation makes life easier for sites, with less technology to navigate so they can better support patients. At its worst, innovation creates new hurdles for sites.

For example, EHR integrations may be a viable trend for the largest sites but are challenging for most sites to adopt. Without the ability to scale a capability, we could create non-standardized solutions that favor some sites while constraining the wider geographic footprint required to reach more patients.

Trials are unrecognizable from 30 years ago
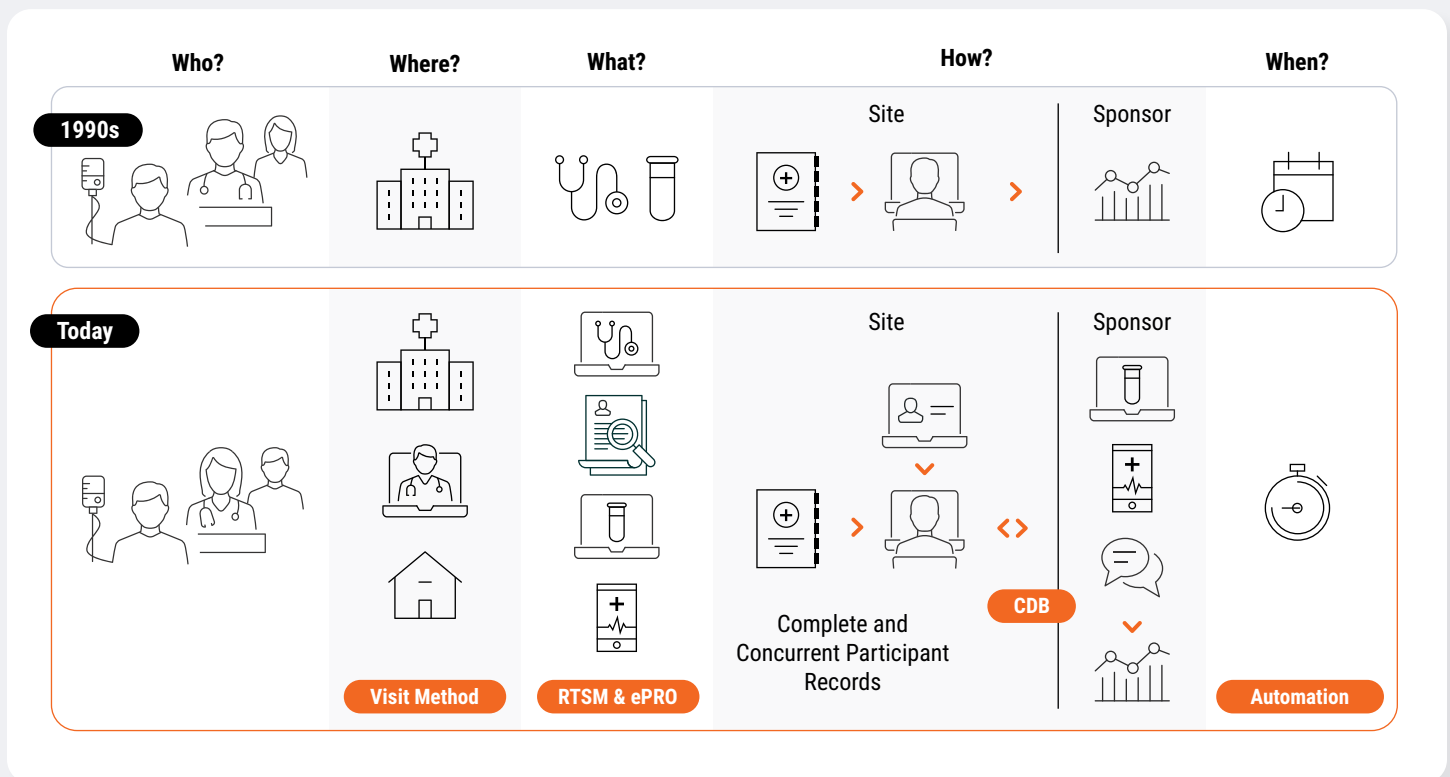
Source: Veeva Systems

Quicker queries are one of many innovations that will simplify *all* sites' experience by reducing the time and effort spent on a time-consuming, manual activity: reviewing and responding to queries. CRAs will no longer have to type common messages and responses: if they see a bad value, they can query that value in one click (similarly, sites would verify, amend, or respond to queried data points in a single click).

One of the ways Alcon, a medical device company, evaluates the site experience is to assess how quickly sites complete data entry: a longer lag could indicate systems and databases are not user-friendly. **A platform-based approach is helping Alcon support site needs.** Leianne Ebert, head of clinical data operations, explains: "This is something we monitor regularly, and last week, our records showed that 45% of our data is entered on the same day as the visit date."

> *Anything that takes away time from patients is a pain point for a site, and anyone who resolves that is helping patient care."*

Vivienne van der Walle
Founder and Medical Director, PT&R

Watch the full episode

Emerging trends: Ones to watch

## 7. Endpoint-driven design enables risk-based approaches

How can risk-based approaches be applied to data management? Clinical research is fundamentally a statistical endeavor, but statisticians are time-poor and typically do most of their work at the backend after data collection. But there are **some indications that statistics invests too much time in activities that do not proportionately improve the quality of the final database** used for analysis: for instance, reviewing, verifying, and querying data that rarely changes.

Endpoint-driven design enables risk-based data management: it limits the amount of data collected and standardized by other functions that do not get used.

If statisticians were involved sooner in data management, they could help define primary and secondary endpoints independently of data management **[Figure 8]**. They would map out the required data, rank it, and identify what's critical (and what isn't). With clearer data cleaning expectations, data management will know which data to focus on and can create queries that advance the study.

Involving stats earlier would improve data management. These teams will question whether endpoint data is missing or unlikely to be used. Over time, endpoint-driven design will lead to less single-use and unnecessary data being collected.

FIGURE 8
Bringing stats to the forefront of data management
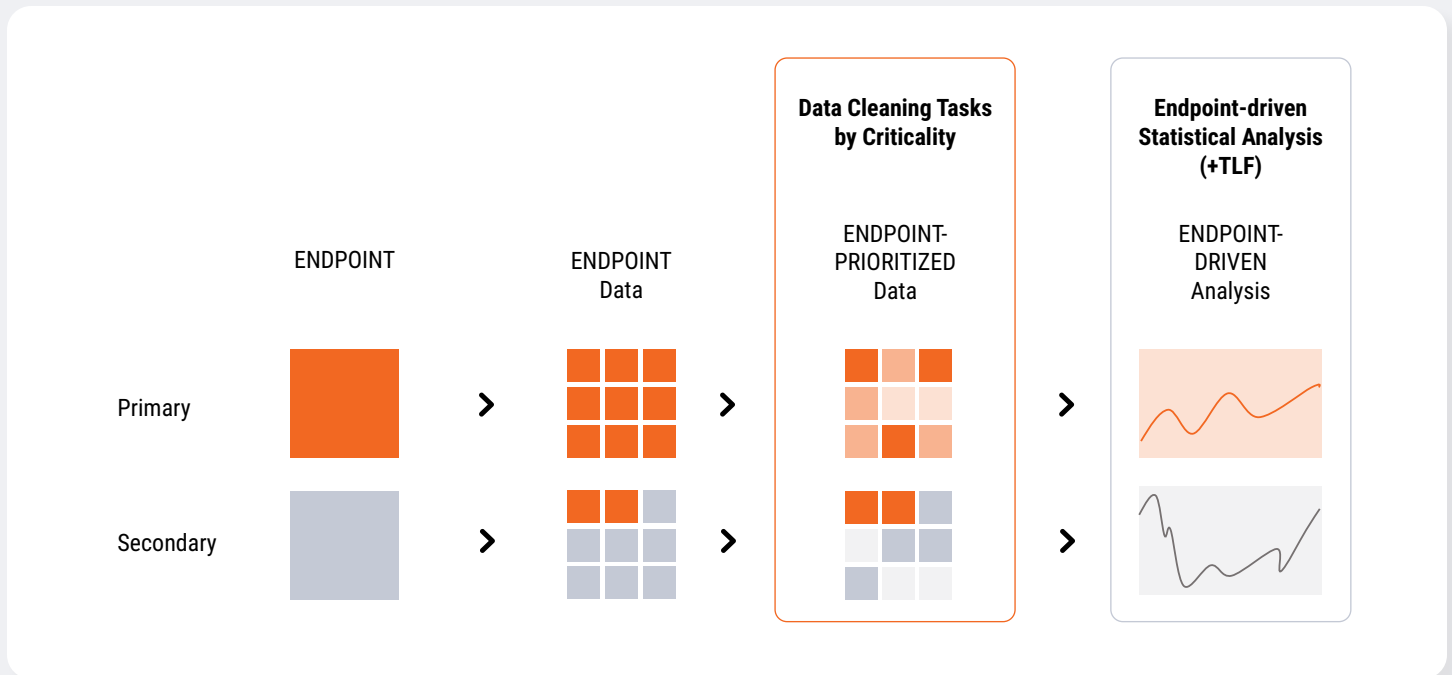
Source: Veeva Systems

8. Boosting data management resilience and future economics

In recent years, **clinical research has continued despite wars, pandemics, and unrelenting macroeconomic pressures.** While the context varies, the scrutiny of value delivered, patient outcomes, and resource efficiency remain consistent.

The new incoming administration in the U.S. could usher a **period of industry uncertainty across several dimensions:** from the direction and leadership of the FDA to the M&A approach favored by the Federal Trade Commission (FTC),

to drug pricing negotiations. In addition, we may see a need for more post-market efficacy studies that bring new data challenges.

Data management will need to become more resilient and grow its strategic contribution to address these risks proactively. As the discipline transitions to clinical data science, it can play more of a strategic role by identifying critical data points, getting involved earlier in the conversation, and educating key decision-makers on the results.

9. Patient choice and diversity move in lockstep

Patient enrollment and retention continue to challenge sites and sponsors, with **trial complexity and patient fatigue** contributing to high dropout rates.

New **FDA draft guidance on Diversity Action Plans** urges sponsors to enroll more patients from underrepresented groups and show the statistical breakdown of participant ethnicity in their trials. This will impact how data is processed and analyzed for submissions, to ensure enough clean data is available to meet statistical endpoints for each demographic subgroup.

Sponsors that give patients more options for onboarding and trial visits will improve their experience and expand the participant pool to previously underserved populations. However, delivering better trials for all means acknowledging the pressure of handling more complex data. Access to connected technologies will become all the more important — from monitoring subgroup enrollment and retention to data management and statistical analysis.

---

10. Sponsors seek increased data ownership and transparency

ICH E6 R2 (and R3 revisions)[5] mandates sponsors to manage risk in a way that reduces the burden on sites, including redundant data review.

Although it doesn't specifically outline a requirement for data ownership, there is growing momentum for sponsors to consolidate and own their data rather than relying on CROs to manage this for them.

Today, there is an observable shift toward:

1. **Fully insourced models**, with outright ownership of the technology and processes
2. **Full data transparency**, with sponsors continuing to outsource, but demanding technology that allows direct access to live data

> *Traditionally, data management was outsourced to our CRO vendor partners. Part of the initiative is to bring all our studies in-house so that our internal teams can start working on it. They can be more hands-on, and we operationalize studies in-house and we are able to take control of our data, and we deliver for our patients with high quality."*
>
> **Head of Clinical Data Engineering**
> Global Biopharmaceutical

Learn more about the trends driving innovation in clinical data.

---

[5] European Medicines Agency, 'ICH E6 (R2) Good clinical practice - Scientific guideline'

Veeva

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**