

The Advantage of Clean Clinical Trial Reference Data

Biopharmaceutical companies continue to increase R&D investment, with over 4,000 global companies spending **\$276 billion** annually. Despite this increased funding, many organizations still rely on inconsistent, siloed systems to manage reference data. These gaps slow daily operations, introduce errors across trial systems, and limit the impact of emerging technologies like AI.

Fragmented data sources once seemed manageable, but the industry’s shift toward higher trial volume and more complex protocols has made clean, connected data essential. A single source of truth for investigator and institution information reduces hidden operational costs, avoids costly study delays, and provides the foundation needed for modern trial execution.

The hidden costs of bad clinical trial reference data

Many sponsors and CROs still manage data manually across multiple systems. For a mid-size biopharma managing

25,000 records, the total hidden annual cost of data cleanup can exceed \$1 million annually. Key drivers include:

01. The ongoing cycle of manual data cleanup for 10,000 impacted records



02. Delays to start-up and trial timelines

Clinical trial procedures and data points collected have **increased by 70%** in the last decade. As trial complexity grows, data errors add up. Inaccurate principal investigator details, outdated site information, or incorrect institutional affiliations can delay site feasibility and selection by weeks.

According to **updated Tufts research**, each lost day is valued at approximately:

- \$500,000 in delayed drug sales
- \$40,000 in added clinical trial expenses

A single source of truth for clinical reference data

A clean data strategy replaces fragmented, reactive cleanup with a continuously curated, global data foundation. Solutions like **Veeva OpenData Clinical** provide a verified directory of investigators and institutions that powers accuracy across all trial systems.

Benefits include:

- **Real-time maintenance and guardrails:** Users can flag missing or incorrect information directly within their systems. Each request routes to the curation team for validation and updates, ensuring data stays clean over time.
- **Operational efficiency and staff focus:** Automated data flows via API eliminate manual reconciliation. Teams can shift their attention from administrative cleanup to strategic activities such as governance, inspection readiness, and AI-enabled decision making.

- **Consistent, verified data at scale:** A global directory maintained by dedicated data curators ensures investigator credentials, affiliations, and site details remain complete and current.

The data foundation for AI-enabled clinical trials

One of the most pressing reasons for clean reference data is the industry's push toward advanced analytics and AI. Many organizations are investing in AI to speed trial operations and improve site selection. However, AI is only as good as the data it consumes. Clinical leaders are already expressing low confidence in the accuracy of their data. A **Harvard Business Review study** shows 97% of data used for business decisions is complete or inaccurate.

Reaping the full benefits of AI tools depends on an organization's ability to maintain consistent, verified data at scale. Transitioning to a single source of data truth provides the standardized foundation. This foundation supports future drug development advancements, ultimately empowering teams to get drugs to patients faster.

Explore how a **single source of truth for clinical trial reference data** can eliminate operational bottlenecks and reduce hidden costs.

