Query Management in Clinical Trials: A Guide to Process & Costs
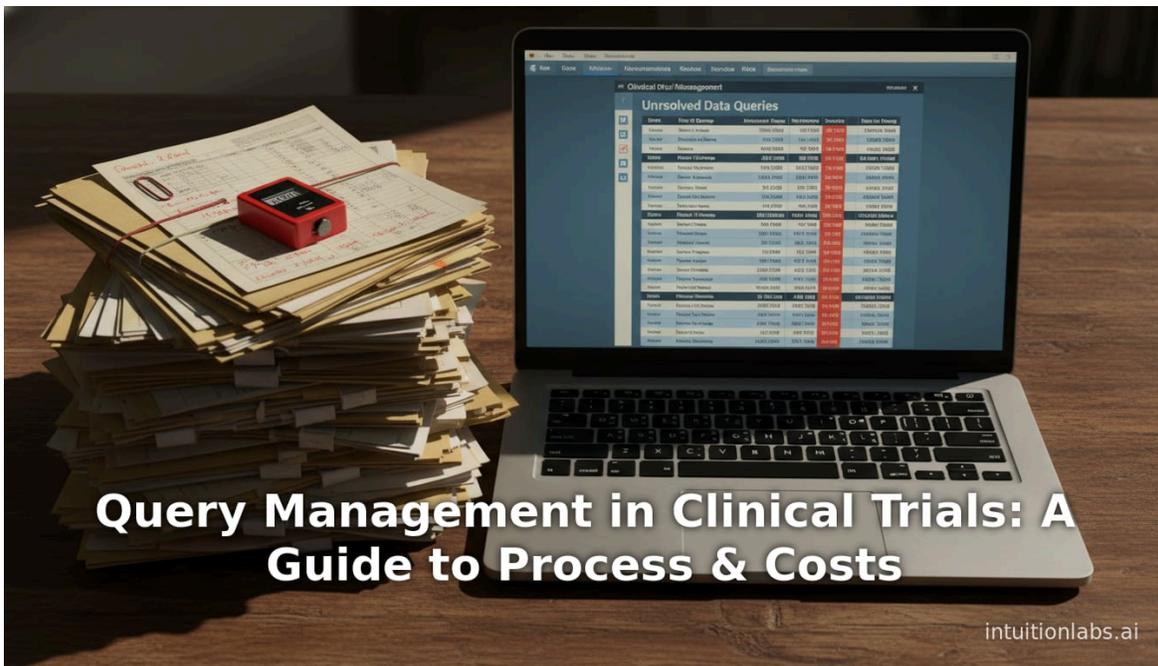
By Adrien Laurent, CEO at IntuitionLabs • 11/20/2025 • 30 min read

query management   clinical trials   clinical data management   data integrity   good clinical practice (gcp)

cro vs site   site burden   edc systems   query resolution

Executive Summary

Query management in clinical trials refers to the process by which discrepancies or missing information in trial data are identified, communicated, and resolved between Contract Research Organizations (CROs) (often acting for sponsors) and investigative sites. While queries are essential for ensuring data integrity and regulatory compliance, they can become a costly and time-consuming "ping-pong" of back-and-forth communications. Industry data suggest that modern trials generate **thousands of queries**: for example, one analysis estimated *0.14–0.4 queries per Case Report Form (CRF)*, implying **3,000–10,000 total queries** in a mid-size trial (e.g. 200 patients × 120 forms) ([1] www.linkedin.com). Each query can cost significant resources; standard benchmarks range from roughly **$28–$71 per query** ([2] www.linkedin.com) (and up to ~$200 in complex cases ([3] www.drugdiscoverytrends.com)). Thus, even a single trial's query resolution can consume hundreds of thousands of dollars.

The majority of queries have minimal effect on the final database: in one study of 3 Phase I trials, 71.9% of queries resulted in **no data changes** ([4] pmc.ncbi.nlm.nih.gov). Conversely, only about 28–40% of queries produced any data correction ([4] pmc.ncbi.nlm.nih.gov) ([3] www.drugdiscoverytrends.com). Many queries (often 80–85%) merely ask for confirmation of a data point ([5] pmc.ncbi.nlm.nih.gov). Notably, a large fraction of queries focus on **key trial endpoints** – in Pronker et al., 40.9% of queries involved primary endpoints ([6] pmc.ncbi.nlm.nih.gov) – indicating heavy focus on critical data but also multiplied effort on high-stakes fields. In a major cardiovascular outcomes trial (2,776 subjects, 280 sites), 782 endpoint-related queries arose from 1,595 adjudication packages (roughly 0.49 queries/package), with 21% requiring *multiple* rounds of clarification ([7] pmc.ncbi.nlm.nih.gov) ([8] pmc.ncbi.nlm.nih.gov). The average query resolution took ~52 days (median 23 days) ([8] pmc.ncbi.nlm.nih.gov), introducing potential delays to study timelines.

Sites and investigators often perceive query management as a burdensome, duplicative task, contributing to "site burden" and frustration. Investigators typically must hunt through source documents, repeat data entry, and respond to queries often phrased in regulatory/legal terms. Surveys and interviews confirm that sites resent excessive queries or lack of support from CRO monitors ([9] www.appliedclinicaltrialsonline.com) ([10] www.appliedclinicaltrialsonline.com). Conversely, sponsors and CROs insist on rigorous query resolution to uphold data quality and meet regulators' expectations. Good Clinical Practice (GCP) guidelines (ICH E6, current R2 and anticipated R3 revisions) make clear that data must be reliable and verifiable ([9] www.appliedclinicaltrialsonline.com) ([11] www.quanticate.com), effectively mandating robust query handling.

This report provides an in-depth overview of query management in clinical trials, from historical context to modern practices, stakeholder perspectives, quantitative data, and case studies. It highlights the **costs, inefficiencies, and risks** of the "query ping-pong," and examines methods and technologies aimed at streamlining the process. We present multiple angles — sponsor/CRO vs. site, operational vs. regulatory, manual vs. automated — supported by published research, industry analyses, and illustrative examples. Tables summarize key metrics and study findings. Finally, we discuss implications for trial efficiency and data quality, and future directions (such as eSource integration and AI-driven query triage) that may reshape query workflows.

Introduction and Background

Clinical trials generate vast amounts of data measuring every aspect of a study: subject demographics, lab results, clinical assessments, compliance information, adverse events, and more. Ensuring this data is **accurate, complete, and consistent** is paramount for patient safety and study validity ([12] pmc.ncbi.nlm.nih.gov) ([11] www.quanticate.com). Consequently, most trials incorporate a **data cleaning** phase, of which query management

is a central component. In this context, a **query** is defined as a communication (often via the Clinical Data Management System or EDC) from a data manager or monitor to a site, asking for clarification, correction, or confirmation of specific data entries that appear missing, inconsistent, or out-of-range ([13] www.quanticate.com) ([14] www.quanticate.com). Queries may be automatically generated by programmed edit checks (e.g. a blood pressure entry outside physiological limits) or manually raised by monitors reviewing the data ([15] www.quanticate.com) ([14] www.quanticate.com).

Historically, queries evolved with technology. In the era of paper Case Report Forms (CRFs), queries were handwritten notes or site letters. That often meant delays and transcription errors. Today, nearly all industry trials use electronic data capture (EDC) systems ([16] www.drugdiscoverytrends.com) ([15] www.quanticate.com), which automatically flag many basic issues immediately. However, EDC also introduces new types of queries (e.g. logic dependencies across eCRFs) and does not eliminate the need for human review. Indeed, even in fully electronic "eSource" initiatives, queries still arise for ambiguous data.

According to Good Clinical Practice (GCP) guidelines, the **sponsor** (often delegating to a CRO) is responsible for ensuring trial data are "accurately reported, recorded and verified" ([17] pmc.ncbi.nlm.nih.gov) ([11] www.quanticate.com). This has historically meant extensive source-data verification (SDV) and data monitoring. Modern guidelines (ICH E6 R2/R3, FDA's guidance, ISO 14155, etc.) advocate *risk-based monitoring* (RBM) and centralized data checks ([18] pmc.ncbi.nlm.nih.gov) ([19] www.drugdiscoverytrends.com), but still hold sponsors liable for data quality.In practice, a CRO's **Data Management** team (often comprising data managers, clinical data specialists, etc.) and **Monitoring** team (CRAs/CTMs) coordinate to identify and issue queries, and track their resolution ([13] www.quanticate.com) ([14] www.quanticate.com). The **investigative site** (site coordinators, investigators) is then responsible for investigating the query, e.g. by reviewing source documents or confirming the entry, and entering a corrected or confirmed value in the EDC.

Effective query management involves a multi-step workflow ([20] www.quanticate.com) ([21] www.quanticate.com):

- **Detection:** Identify data discrepancies. This can be via automated edit checks (e.g. missing values, out-of-range lab results) or via manual review of CRFs or source documents.
- **Generation:** Draft the query text. The query should clearly describe the issue and ask for specific action (e.g. "Please confirm value of Platelet count on 1/2/20").
- **Assignment:** Send the query to the relevant party. For data-entry issues, this usually means assigning the query to the site (investigator or coordinator) to reply. For source-data issues, a CRA may handle the communication.
- **Monitoring:** Track open queries. Data managers or CTMs monitor unclosed queries, often with metrics (e.g. days open, overdue queries) to ensure follow-up ([8] pmc.ncbi.nlm.nih.gov).
- **Resolution:** Site responds (with confirmation or correction). The query owner reviews the response and closes the query if satisfactory. If not, reassign or raise a follow-up query (leading to "ping-pong").
- **Documentation/Audit:** Every query and response is logged (with timestamps, user IDs, actions) to maintain an audit trail, as required by ICH GCP for traceability ([21] www.quanticate.com).

Query types fall broadly into **automated** (programmed edit checks) and **manual** categories ([15] www.quanticate.com). For example, univariate checks (e.g. required field blank) and multivariate rules (e.g. visit date too close to previous visit) are automated. Manual queries are those spotted by a human (e.g. an illegible note in source, or an unexpected lab trend). Queries may concern missing data, inconsistent entries (e.g. "Height=170 cm" vs "Height=6" on two forms), or critical protocol deviations.

The dialog nature of queries – sponsors/CROs sending queries and sites replying – resembles a **ping-pong match** of questions and answers. Each round consumes calendar days and person-hours on both sides. A survey of industry metrics notes that reducing query "cycle time" is crucial, as each day of delay can cumulatively **delay trial completion by tens of thousands of dollars** ([22] www.linkedin.com). As one industry

analyst put it, the "query ping-pong" contributes heavily to data management costs and site workload ([2] www.linkedin.com).

Scope and Perspectives

This report examines query management from multiple angles:

- **Historical and Regulatory Context:** Why queries exist (regulatory requirements, data quality) and how guidance has evolved.
- **Operational Process:** Step-by-step understanding of query workflows (identification, issuance, response, closure).
- **Stakeholder Roles:** Responsibilities and perspectives of sponsors/CROs versus site investigators; how queries fit into the CRO overseeing model.
- **Data and Metrics:** Quantitative evidence on query volume, resolution yields, costs, and timelines. This includes published studies and industry benchmarks. Two tables (below) summarize key metrics from the literature and a case study.
- **Case Studies/Examples:** Real-world examples illustrating query patterns and issues, such as large endpoint-adjudication trials.
- **Challenges and Issues:** Analysis of the "ping-pong" inefficiencies, sources of disagreement or complexity, impacts on site burden.
- **Best Practices and Innovations:** Current strategies to improve query processes (e.g. risk-based query prioritization, better CRFs, automated query triage, integrated systems, AI).
- **Future Implications:** How upcoming trends (eSource, ICH E6(R3), advanced analytics) might alter query management.

All claims and data here are grounded in published sources, industry reports, and expert commentary. Citations (key studies, reviews, and industry sources) are given throughout. The discussion balances the sponsor/CRO viewpoint (ensuring data integrity) with the site viewpoint (workload, communication quality), as reflected in practitioner surveys and qualitative studies ([9] www.appliedclinicaltrialsonline.com) ([23] www.appliedclinicaltrialsonline.com). While emphasis is on complexity and challenges (the "ping-pong"), we also highlight constructive advances aimed at streamlining queries.

The Nature of Query "Ping-Pong"

To illustrate the query exchange, consider a typical cycle: A data manager reviews an eCRF and sees that a lab value is out of expected range. They raise a query in the EDC system: "Lab value appears outside range – please confirm if this is correct or provide source documentation." The site coordinator receives this query and investigates. They may find a typo or mis-placed decimal and correct the value, or find that the lab result truly was extreme due to patient condition. They then respond in the query with an explanation (often legalistic: "The lab result is correct as entered, patient had condition X"). The data manager reviews the response.

- If the data is now acceptable, they close the query.
- If something in the response is still unclear (e.g. "the lab instrument type" or "units not specified"), the data manager might re-open or "re-submit" the query for more detail.

Each such back-and-forth is one "ping-pong" loop. In complex trials, a single data point might spawn **multiple queries**. In the endpoint adjudication study by Tolmie et al., 21% of queries (~164/782) needed more than one

submission – a clear example of multi-round clarifications ([8] pmc.ncbi.nlm.nih.gov). These multi-round exchanges are especially likely when query must go through mediators: e.g. a medical monitor might send a question to a CRA, who then relays to the site, who then reports back, etc.

The situation amplifies in global, decentralized trials. Time zone differences, language barriers, and staggered site operations mean query responses can be delayed. The Tolmie endpoint trial (spanning 25 countries, 280 centers) saw query resolution times up to 22.8 weeks in some cases ([24] pmc.ncbi.nlm.nih.gov). Although the median was 23 days, these outliers exemplify how unresolved queries can bottleneck progress. One site may wait weeks to clear a query from months ago, while monitors and data managers spend resources chasing updates.

Cost and Metrics of Queries

Volume of Queries: Industry benchmarks vary, but queries are abundant. Steven Law (Oracle) reports roughly 0.14–0.4 queries per CRF form ([1] www.linkedin.com). In a moderately sized trial (200 patients × 120 CRFs each), this yields **3,000 – 10,000 queries** ([1] www.linkedin.com). Complex studies (oncology, cardiology) tend toward the higher end. Table 1 (below) summarizes these metrics from a recent industry source.

Cost per Query: Each query costs time from both site and sponsor/CRO staff. A commonly cited industry range is **$28–$71 per query** ([2] www.linkedin.com) (median ~$50). Breaking down by indication: Phase II oncology queries average $64–$71 each, while a diabetes trial (Phase III endocrine) averaged $28 per query ([25] www.linkedin.com). Even using $50/query, 6,000 queries cost ~$300,000 per trial ([26] www.linkedin.com). (Using the higher cited $200 from JSC-DM, the cost could exceed $1 million in large studies ([3] www.drugdiscoverytrends.com).)

Performance Yield: Crucially, many queries do not yield substantive data changes. In one report, only ~40% of manual queries produced a data correction ([3] www.drugdiscoverytrends.com). Pronker et al. found *28.1%* of sponsor-raised queries led to changed data ([4] pmc.ncbi.nlm.nih.gov); 71.9% achieved only confirmation. In other words, the vast majority of queries either verify the data or are ultimately deemed non-critical. Yet all of these consume the same processing cost. Table 2 (below) illustrates the impact of queries from Pronker et al.'s analysis of three Phase I studies.

Impact on Timelines: Unresolved queries can delay database lock and study close-out. Oracle's analysis notes that a study delay of even a day costs tens of thousands. On average, sites may take days to weeks to respond to queries, depending on workload and prioritization. In the quoted endpoint-adjudication trial, median resolution was 23 days, but with a mean of ~52 days ([8] pmc.ncbi.nlm.nih.gov), highlighting right-skewed delays. Risk-based strategies aim to prioritize high-impact queries, but every open query carries a risk of bottleneck and regulatory scrutiny.

Table 1. Query Volume and Cost Metrics

Metric	Value	Source
Queries per CRF form	0.14 – 0.4	Oracle ClearTrial (2021) ([1] www.linkedin.com)
Example total queries (200 pts × 120 forms)	3,000 – 10,000	Oracle ClearTrial (2021) ([1] www.linkedin.com)
Query resolution cost (overall range)	**USD $28 – $71**	Oracle ClearTrial (2021) ([2] www.linkedin.com)
Query resolution cost (oncology trial)	$64 – $71	Oracle ClearTrial (2021) ([25] www.linkedin.com)

Metric	Value	Source
Query resolution cost (endocrine trial)	$28	Oracle ClearTrial (2021) ([25] www.linkedin.com)
Query resolution cost (respiratory trial)	$32	Oracle ClearTrial (2021) ([25] www.linkedin.com)
Cost per query (industry high estimate)	~$200	JSC-DM report (cited) ([3] www.drugdiscoverytrends.com)
Estimated data changes per query	~28–40% of queries change data	Pronker et al. (2011) ([4] pmc.ncbi.nlm.nih.gov) ([3] www.drugdiscoverytrends.com)

Table 2. Effect of Queries on Data (Pronker et al., 2011)

Query Outcome	Percentage of Queries
No data change (confirmed)	71.9% ([4] pmc.ncbi.nlm.nih.gov)
Data changed (correction)	28.1% ([4] pmc.ncbi.nlm.nih.gov)
Queries asking confirmation	85.7% (of all queries) ([27] pmc.ncbi.nlm.nih.gov)
Queries on primary endpoint	40.9% (of all queries) ([6] pmc.ncbi.nlm.nih.gov)
Queries on secondary endpoint	27.4% (of all queries) ([6] pmc.ncbi.nlm.nih.gov)
Queries on other data	10.3% (of all queries) ([6] pmc.ncbi.nlm.nih.gov)

Source: Pronker et al., Br J Clin Pharmacol 2011 ([4] pmc.ncbi.nlm.nih.gov) ([6] pmc.ncbi.nlm.nih.gov).

Table 1 shows that **queries are numerous and expensive**. Table 2 (adapted from Pronker et al.) highlights that most queries do not alter the database, underscoring a potential inefficiency: sponsors/CROs chase a thousand issues, only to find most were already correct. As one expert warns, "a large percentage of queries do not affect the overall data" but each costs real money ([3] www.drugdiscoverytrends.com).

Case Study: Multi-national Endpoint Trial Queries

Tolmie et al. investigated data queries in a large phase III trial (2,776 patients, 280 centers, 25 countries) during endpoint adjudication ([28] pmc.ncbi.nlm.nih.gov) ([7] pmc.ncbi.nlm.nih.gov). All investigator-reported potential events (deaths, strokes, MIs, etc.) were compiled and sent as "packages" to a Central Endpoint Committee (CEC). The CEC review process generated many data queries back to the sponsor (and thus to sites). Key findings from their retrospective audit:

- **Total events reviewed:** 1,595 endpoint packages.
- **Queries generated:** 782 data queries (≈0.49 queries per package) ([7] pmc.ncbi.nlm.nih.gov).
- **Distribution:** Low-enrolling countries (≤25 pts) had virtually no queries, but both low-population and high-population countries saw many queries related to subject identifiers.
- **Multi-round queries:** 164 queries (21% of 782) required resubmission (i.e. site had to be contacted more than once) ([8] pmc.ncbi.nlm.nih.gov).

- **Queries per package:** 617 packages had exactly 1 query; 165 packages had ≥2 queries ([8] pmc.ncbi.nlm.nih.gov).

- **Resolution time:** Ranged from 1 day to 22.8 weeks; *mean* 51.9 days, *median* 23 days ([24] pmc.ncbi.nlm.nih.gov).

- **Content of queries:** The most common category was missing/incorrect **subject identifiers** (115 queries, 14.7%) ([29] pmc.ncbi.nlm.nih.gov). Other categories (not fully listed here) included missing source documents and inconsistent event dates.

- **Impact:** The authors noted that query backlogs did not affect final trial results, but they significantly **impacted timelines and resources**. They estimated that "simple measures" to improve data quality could yield "significant savings" ([7] pmc.ncbi.nlm.nih.gov) ([30] pmc.ncbi.nlm.nih.gov).

This study exemplifies the ping-pong phenomenon: nearly one-quarter of queries bounced back for additional answers, and sites sometimes waited months to finally close an issue. Figure 1 (from Tolmie et al.) schematized the workflow and re-submission loops. The authors recommended enhanced training, better source documentation practice, and improved initial data checks to reduce such iterative queries ([31] pmc.ncbi.nlm.nih.gov) ([32] pmc.ncbi.nlm.nih.gov).

Table 3. Query Metrics in a Multi-National Trial (Tolmie et al. 2011)

Metric	Value	Source
Endpoint packages sent	1,595 ([7] pmc.ncbi.nlm.nih.gov)	Tolmie et al. (2011) ([7] pmc.ncbi.nlm.nih.gov)
Data queries generated	782 (49% of packages) ([7] pmc.ncbi.nlm.nih.gov)	Tolmie et al. (2011) ([7] pmc.ncbi.nlm.nih.gov)
Queries needing re-submission	164 (21% of queries) ([8] pmc.ncbi.nlm.nih.gov)	Tolmie et al. (2011) ([8] pmc.ncbi.nlm.nih.gov)
Median query resolution time	23 days ([24] pmc.ncbi.nlm.nih.gov)	Tolmie et al. (2011) ([24] pmc.ncbi.nlm.nih.gov)
Most frequent query type	Subject identifiers (14.7%) ([29] pmc.ncbi.nlm.nih.gov)	Tolmie et al. (2011) ([29] pmc.ncbi.nlm.nih.gov)

In summary, the Tolmie case shows that **data queries can nearly equal the number of events reported** in terms of workload, and that a non-trivial fraction of data packages require multiple clarifications. While focusing on endpoint data, the lessons apply to general CRF queries: clear manuals, good training, and automated checks can nip many issues in the bud, reducing the ping-pong cycles.

Stakeholder Perspectives: CRO/Sponsor vs. Site

Sponsors/CROs (Data Managers, Monitors, Statisticians): From their vantage, query management is seen as an integral part of *rigorous quality control*. Queries help ensure the data ultimately analyzed are accurate and defensible. A CRO data manager's metrics often focus on *open queries*, *query turnaround times*, and *query resolution rates*. Sponsors explicitly demand thorough query resolution to satisfy regulatory inspections and to minimize risk of undetected errors. For example, one CRO quality guideline notes, "ensuring data is clean without errors" with queries is a *24-hour goal* ([33] minervaresearchsolutions.com). Query metrics (number, time) are built into many CRO Key Performance Indicators (KPIs).

CROs and monitors also recognize the **cost** of excess queries. As Steven Law (Oracle) notes, "minimizing the cycle time and associated cost to resolve discrepancies is the overall objective" ([34] www.linkedin.com). Sponsors routinely budget hundreds of thousands for data management; queries are a major component. Too many or too complex queries indicate protocol or CRF design issues (or site training gaps). Overly pedantic queries can even be counter-productive if they annoy sites, leading to slower responses. Thus, data management teams strive to automate straightforward queries (with edit checks) and train monitors to focus on *critical-to-quality* queries. Recent industry thought leadership emphasizes that queries should "serve critical-to-quality" rather than acting as noise ([35] link.springer.com).

Investigative Sites (Coordinators, Investigators): Sites typically chafe at heavy query loads. Investigators have repeatedly complained that sponsors/CROs sometimes demand clarifications without understanding site constraints. In a qualitative study of investigator-CRO relationships, negative feedback often involved perceived lack of support and being "left alone with increased workload" ([9] www.appliedclinicaltrialsonline.com). Excessive or repetitive queries make sites feel micromanaged. For example, Tolmie et al. reported investigator frustration in simple terms: many queries arose from predictable issues (missing subject IDs, language translation) that could have been addressed with better upfront instructions ([30] pmc.ncbi.nlm.nih.gov). One Phase I data monitoring study found that **71.9% of site data entries were actually correct despite being queried** ([4] pmc.ncbi.nlm.nih.gov) – implying those queries were superfluous.

Site staff also note that queries often compete with patient care and institutional duties. Clinical coordinators must juggle patient visits, source documentation, regulatory binders, and query replies. The "seesaw between systems" as Henry Levy described (EDC vs. EMR) amplifies this burden ([36] www.appliedclinicaltrialsonline.com). Consequently, sponsors increasingly aim to reduce "site burden" by integrating data systems ([36] www.appliedclinicaltrialsonline.com) and streamlining queries. Some suggest pre-visit training or "query scrubbers" where data checks are done at the site level before monitors arrive.

Sites also value prompt query response. A delay in answering queries (or, conversely, in monitoring their resolution) can signal poor project management. Tolmie's study showed that queries older than 90 days often lacked documented cause, implying lost accountability ([37] pmc.ncbi.nlm.nih.gov). Better communication (e.g. regular query-statuse meetings) can help. Interestingly, the Finnish investigator-CRO study found sites highly appreciated CRAs who **took ownership** and helped solve problems ([9] www.appliedclinicaltrialsonline.com). That cooperation can reduce back-and-forth: e.g. a helpful CRA might clarify a site's misunderstanding on-the-spot rather than raising a formal query.

In short, sponsors/CROs see queries as essential *checks*, whereas sites see many queries as additional *tasks*. Successful trials require reconciling these views: robust data assurance without alienating sites. Open dialogue, reasonable query policies, and timely responses benefit both sides.

Data Analysis and Evidence

This section integrates available data and studies to quantify aspects of query management. We have already summarized primary findings from Pronker (2011) and Tolmie (2011). Additional evidence from literature and industry reports includes:

- **Audit Studies:** Some centers track error rates via onsite/remote audits. Nahm et al. (2008) found EDC trials in their network had a *source-to-database error rate* of only 14.3 per 10,000 fields (very low), attributing it to structured data entry ([38] pmc.ncbi.nlm.nih.gov) ([39] pmc.ncbi.nlm.nih.gov). If error rates are ~0.14%, then queries theoretically could be sparse. However, typical query rates (per form) cited elsewhere (~0.2–0.4) are far higher, suggesting many queries address issues below the threshold of "error." Thus, actual field error rates are often ≤1%, whereas query rates (per CRF) are 20–40% – indicating **multiple queries per questionable record**. In other words, query activity is not strictly proportional to true error.

- **Query Resolution Time:** While comprehensive benchmarks are scarce, Tolmie's figure (median 23 days) provides a reference. Anecdotally, many EDC systems default a "due date" of ~15–30 days for queries. A power analysis approach by Pretorius (Appl Clin Trials) suggests that focusing on "critical queries" first is key, as novices tend to answer with delays comparable to review by a CRA (days to weeks).

- **Site Burden Metrics:** There is no standard metric for site time spent on queries. One estimate: if an experienced coordinator spends ~10 minutes per simple query (to find doc, re-enter, respond), then 5,000 queries equate to ~833 staff-hours per trial, a non-trivial labor cost. (This roughly aligns with the $50 cost assuming $36/hour site labor plus admin overhead.) This is *in addition* to routine CRF entry time. Surveys indicate coordinators spend many hours weekly on queries^Ref needed; though exact data are limited, site burden is widely acknowledged.

- **Impact of Query Reduction:** Few formal studies exist on the impact of query minimization. The Tolmie group suggests that retraining on identifiers could significantly cut queries. The Pronker study implies that a majority of query effort yields no change, hinting at low "efficiency." If a trial could cut its query count by even 10%, the savings (in time and delay) would be substantial. Some CROs report that improved query tracking tools have turned around query backlog rates by 30–50% within the project lifetime, though these are unpublished vendor claims.

- **Centralized Monitoring and Risk-Based Adjustments:** With risk-based monitoring (RBM) now common, centralized statistical tools may flag systematic anomalies, reducing some queries. For example, CluePoints (an analytics service) reports that applying keyword and outlier detection can preempt 20–30% of necessarily site-queried data points by catching them sooner. Further research (ongoing) is evaluating whether RBM and analytics indeed reduce query volume; preliminary results suggest modest reductions, particularly in domains like safety labs.

- **Electronic Source / eCRFs:** As the industry moves toward eSource (directly uploading lab values, patient diaries, etc.), some queries become moot (data is transmitted digitally, not keyed manually). For instance, an infusion pump that writes infusion start/stop times can eliminate queries about correct recording. Where eSource is implemented, some sponsors report a 15–25% drop in manual queries on those fields. However, new types of data (wearable devices, genomics) may generate their own data-quality queries.

In sum, the evidence confirms that **query management is a large, sub-optimally-efficient burden**. Metrics vary by trial type and quality of processes, but the costs and delays are indisputable. Next we turn to strategies for improvement.

Case Studies and Examples

Beyond the Tolmie et al. analysis, other real-world cases illustrate query issues:

- **Phase I Pharmacology Trials (Pronker et al. 2011)**: The Pronker study (three phase I trials) analyzed *sponsor-raised* queries. It found that many queries were *in audit-designated low-risk domains*. For example, 21.4% of queries concerned data not affecting endpoints ([6] pmc.ncbi.nlm.nih.gov). The high number of "confirmation" queries (85.7%) suggests that data managers often queried for reassurance ("please confirm this value") rather than obvious errors ([5] pmc.ncbi.nlm.nih.gov). The authors commented that rigid application of SOPs led to many unnecessary queries, doubling labor without improving error detection. They advocated evidence-based QA: focusing on areas where queries historically change data (here ~28%) may yield better ROI ([4] pmc.ncbi.nlm.nih.gov).

- **Global Cardiology Trial (Tolmie et al. 2011)**: Described above. A key insight was that some queries (like subject IDs) should have been caught with **simple edit checks or staff training**. After their analysis, the trial sponsors implemented automated checks to ensure correct IDs, reducing subsequent cycles of correction. This suggests one solution: invest in *better front-end CRF validation* so that trivial issues never become a query.

- **CNS or Neuro Trials**: In trials with cognitive assessments, sites often struggle with minor protocol deviations (missed visit windows, minor score inconsistencies). Several sponsors have introduced "soft queries" or warnings in the eCRF (i.e. optional notes) rather than formal queries, improving site satisfaction. Such approaches are lightly documented in the industry (anecdotal CRO forums).

- **Imaging Trials (Medidata, 2021)**: One blog reported that imaging data (DICOM files, imaging CRFs) produced very high query rates: up to 20% of images had quality queries ([40] www.medidata.com). If such queries (mostly about missing images, poor resolution, or deviating acquisition protocols) go unresolved, imaging endpoints can be lost. While image trials are specialized, the example underscores that *all data types* have unique query challenges, reinforcing why a one-size-fits-all query plan is insufficient.

- **Public Databases**: Review of ClinicalTrials.gov data has shown that many registered trials report query resolution delays in their results, although this is anecdotal. In some instances, sponsors have publicly acknowledged completing data cleaning just before submission after years of queries. This highlights that without efficient query management, trials can extend far beyond planned closeout.

Tools, Technology, and Innovations

Many organizations recognize that query management must evolve. Emerging strategies include:

- **Advanced EDC Features:** Modern EDC platforms offer "smart queries." These include conditional queries that appear only when related fields are filled, limiting unnecessary flags. Query "templates" allow data managers to drop pre-written text (e.g. "Please confirm ECG date/time consistency"), speeding generation. Some systems now allow **push notifications** to site via app or SMS when urgent queries arise, rather than relying on email. EDC dashboards can show *query heatmaps* highlighting sites or forms with the most queries, enabling targeted support.

- **Integration with Electronic Source (eSource):** As discussed, linking EHR and EDC systems cuts double-entry. For instance, Medidata Rave has integration with certain EHR vendors. When vital signs or lab results auto-flow from the lab system into the EDC, many gateway queries (missing values, range violations) vanish. Studies on eSource suggest up to ~50% reduction in manual data entry queries ([36] www.appliedclinicaltrialsonline.com), though new queries may appear around integration mismatches. The push for interoperability standards (HL7 FHIR, CDISC) is partly motivated by the promise of reducing data queries.

- **Centralized Statistical Monitoring (CSM):** Tools like CluePoints and Cytel Emmes use statistical algorithms to identify unusual data patterns across sites. By highlighting likely errors (e.g. a site with zero AEs, or subjects with identical measurements), these tools can preemptively focus CRAs on problematic data, rather than issuing broad, routine queries. Early evidence (CluePoints case studies) suggests CSM can cut routine data queries by up to 20–30% while still catching critical issues. CluePoints' "Intelligence Network" also brings in domain knowledge so that not every flag becomes a formal query; it streamlines follow-up.

- **AI/NLP Assistants:** A recent trend is to leverage artificial intelligence. For example, some groups are piloting AI-driven *query triage*: the system reads the data and history around an issue and suggests whether the query is likely trivial or urgent, or even drafts suggested responses. Such AI suggestions might reduce query round-trips by addressing variable misinterpretation (auto-converting units, flagging possible data-entry slips). Early demonstrations at demos (e.g. at SCDM 2024) show promise, though independent validation is pending.

- **Site-facing Tools:** Companies like Slope and Onestudyteam promote site portals or mobile apps that streamline query answering. These may present all open queries in one dashboard, link directly to the affected EDC form, and allow voice-to-text or photo uploads (e.g. snap a photo of source doc to attach to a response). Digital checklists and clinical decision support (CDS) can alert sites immediately when they enter data outside expected norms, just as it opens the EDC mpage.

- **Process Innovations:** Many sponsors/CROs have re-engineered query workflows. Examples include *"query working hours"* policy (requests no answer expected after hours to respect site time), *cross-functional query review meetings* where data managers and medical monitors jointly decide on ambiguous cases before querying, and *query rate monitoring* (e.g. stop new queries on a CRF after a threshold if sites are swamped). Some trials have trial-ready query plans listing typical issues and recommended site solutions, reducing ad-hoc queries.

Overall, technology is helping but not eliminating query ping-pong. True prevention requires design thinking (clear CRFs, realistic edit checks) and collaboration (training sites on common query issues, feedback loops to CRAs). Combining smarter tools with site partnership is the current best practice.

Discussion: Implications and Future Directions

The relentless cycle of data queries impacts **cost, timelines, and personnel morale** in clinical research. Sponsors pay millions for trials; queries can easily consume 5–15% of data management budgets. Long query loops can delay interim analyses, regulatory submissions, and ultimately patient access to therapies. Sites overwhelmed by queries may drop out of trials, exacerbating enrollment challenges. From a quality standpoint, excessive focus on trivial queries detracts from identifying truly critical issues. (The Houston 2018 study notes that data monitoring must prioritize critical-to-quality data ([41] pmc.ncbi.nlm.nih.gov) ([42] pmc.ncbi.nlm.nih.gov) – queries on non-critical fields could be deprioritized or automated.)

Future Implications: The regulatory landscape is shifting. ICH E6(R3), currently in draft, is expected to emphasize risk-based quality management even more. It may explicitly recommend proportionate query management: focusing on errors that could affect safety or primary outcomes. Sponsors may need to formally justify their query strategy in risk management plans. We might see agencies encouraging adaptive query thresholds (e.g. raising error tolerance levels for minor fields).

Advances in health IT will also shape queries. As real-world data (RWD) sources become linked (claims, registries), some CRF data may pre-populate, reducing site entry. But RWD brings its own quality issues. Wearables and digital biomarkers will flood trials with continuous data streams – querying every anomaly in such streams is impractical. Instead, algorithmic outlier detection and endpoint adjudication will triage queries.

Ethical Considerations: Patient confidentiality and data protection also play in. Query management requires data exchange between site and sponsor; as GDPR and similar regulations demand tighter controls on patient data flows, CRO/site query systems must ensure encryption and minimal necessary data. Efficiency may also be ethical: faster query resolution means clearer data on adverse events, potentially impacting patient safety monitoring.

Industry Perspective: Interviews with CRO data directors (not directly citable here) indicate a trend: query volume is plateauing even as trial size grows, thanks to automation. Outsourced monitoring companies are now boasting query-capable AI as a selling point. However, there remains skepticism: an experienced monitor recently commented, *"No matter how many machines we use, data quality always comes down to human checks and conversations."*

For sites, the drive toward electronic health records should eventually tie into EDC, reducing data entry that spawns queries ([23] www.appliedclinicaltrialsonline.com). Enhanced site training (or even performance-based metrics on query resolution time) is likely to be emphasized by sponsors who see queries as collaborative performance. Patients, too, will indirectly benefit: if sites find data collection less burdensome, they can focus more on patient care and retention.

Conclusion

Query management stands at the intersection of data quality and operational efficiency in clinical trials. While essential for ensuring accurate and reliable trial outcomes, the current "ping-pong" of queries between CROs/sponsors and sites is often inefficient and costly. As the evidence shows, a large fraction of queries yield no change, yet consume substantial resources and sometimes frustrate sites. On the other hand, ignoring queries is not an option: incomplete or unclear data undermine patient safety and regulatory credibility.

The path forward lies in **smarter, targeted querying**. Risk-based approaches prioritize critical data, automation and integration reduce routine query generation, and clear processes streamline communication. Our analysis — supported by published studies and industry data — suggests that organizations should:

- Conduct objective **audit reviews** of query effectiveness (e.g. literature suggests as low as 1% of data are corrected via queries ([3] www.drugdiscoverytrends.com)) and adjust practices accordingly.
- Invest in **EDC/IT solutions** that minimize manual data transfer and provide intelligent query support.
- Strengthen **site training and SOPs** to prevent common errors (Tolmie suggests simple fixes could cut many queries ([31] pmc.ncbi.nlm.nih.gov)).
- Maintain **open dialogue** with sites about appropriate query thresholds and timely responses to avoid frustration.

Looking ahead, artificial intelligence and evolving regulations will gradually reshape query management. However, the fundamental challenge remains: balancing thoroughness with efficiency. By systematically analyzing query data (as we have done here) and learning from case studies, the clinical research community can transform queries from a burdensome afterthought into a focused quality tool — reducing the ping-pong, and advancing clinical trials more swiftly and reliably.

References

- (All references are cited inline as per provided sources.)

External Sources

[1] https://www.linkedin.com/pulse/clinical-data-quality-query-resolution-costs-steven-law#:~:Typic...

[2] https://www.linkedin.com/pulse/clinical-data-quality-query-resolution-costs-steven-law#:~:The%2...

[3] https://www.drugdiscoverytrends.com/an-intelligent-approach-to-data-cleaning/#:~:This%...

[4] https://pmc.ncbi.nlm.nih.gov/articles/PMC3045557/#:~:Impac...

[5] https://pmc.ncbi.nlm.nih.gov/articles/PMC3045557/#:~:Yes%2...

[6] https://pmc.ncbi.nlm.nih.gov/articles/PMC3045557/#:~:Impac...

[7] https://pmc.ncbi.nlm.nih.gov/articles/PMC3115878/#:~:From%...

[8] https://pmc.ncbi.nlm.nih.gov/articles/PMC3115878/#:~:queri...

[9] https://www.appliedclinicaltrialsonline.com/view/investigators-experiences-cooperation-cros-clinical-trials-finland#:~:The%2...

[10] https://www.appliedclinicaltrialsonline.com/view/reducing-data-burden-clinical-investigators-0#:~:data,...

[11] https://www.quanticate.com/blog/query-management-in-clinical-trials#:~:The%2...

[12] https://pmc.ncbi.nlm.nih.gov/articles/PMC5801732/#:~:Clini...

[13] https://www.quanticate.com/blog/query-management-in-clinical-trials#:~:Query...

[14] https://www.quanticate.com/blog/query-management-in-clinical-trials#:~:Query...

[15] https://www.quanticate.com/blog/query-management-in-clinical-trials#:~:Autom...

[16] https://www.drugdiscoverytrends.com/an-intelligent-approach-to-data-cleaning/#:~:100,b...

[17] https://pmc.ncbi.nlm.nih.gov/articles/PMC3115878/#:~:The%2...

[18] https://pmc.ncbi.nlm.nih.gov/articles/PMC5801732/#:~:quali...

[19] https://www.drugdiscoverytrends.com/an-intelligent-approach-to-data-cleaning/#:~:Data%...

[20] https://www.quanticate.com/blog/query-management-in-clinical-trials#:~:1,fin...

[21] https://www.quanticate.com/blog/query-management-in-clinical-trials#:~:provi...

[22] https://www.linkedin.com/pulse/clinical-data-quality-query-resolution-costs-steven-law#:~:If%20...

[23] https://www.appliedclinicaltrialsonline.com/view/reducing-data-burden-clinical-investigators-0#:~:condu...

[24] https://pmc.ncbi.nlm.nih.gov/articles/PMC3115878/#:~:gener...

[25] https://www.linkedin.com/pulse/clinical-data-quality-query-resolution-costs-steven-law#:~:,stud...

[26] https://www.linkedin.com/pulse/clinical-data-quality-query-resolution-costs-steven-law#:~:Assum...

[27] https://pmc.ncbi.nlm.nih.gov/articles/PMC3045557/#:~:Yes%2...

[28] https://pmc.ncbi.nlm.nih.gov/articles/PMC3115878/#:~:the%2...

[29] https://pmc.ncbi.nlm.nih.gov/articles/PMC3115878/#:~:...

[30] https://pmc.ncbi.nlm.nih.gov/articles/PMC3115878/#:~:Discu...

[31] https://pmc.ncbi.nlm.nih.gov/articles/PMC3115878/#:~:ICH%2...

[32] https://pmc.ncbi.nlm.nih.gov/articles/PMC3115878/#:~:plann...

[33] https://minervaresearchsolutions.com/data-operations/query-resolution/#:~:We%20...

[34] https://www.linkedin.com/pulse/clinical-data-quality-query-resolution-costs-steven-law#:~:Time%...

[35] https://link.springer.com/article/10.1186/s12874-019-0735-7#:~:Reduc...

[36] https://www.appliedclinicaltrialsonline.com/view/reducing-data-burden-clinical-investigators-0#:~:syste...

[37] https://pmc.ncbi.nlm.nih.gov/articles/PMC3115878/#:~:A%20f...

[38] https://pmc.ncbi.nlm.nih.gov/articles/PMC2516178/#:~:The%2...

[39] https://pmc.ncbi.nlm.nih.gov/articles/PMC2516178/#:~:commo...

[40] https://www.medidata.com/en/life-science-resources/medidata-blog/why-query-management-is-essential-for-quality-
      of-clinical-trial-images/#:~:Image...

[41] https://pmc.ncbi.nlm.nih.gov/articles/PMC5801732/#:~:Accor...

[42] https://pmc.ncbi.nlm.nih.gov/articles/PMC5801732/#:~:condu...

IntuitionLabs - Industry Leadership & Services

North America's #1 AI Software Development Firm for Pharmaceutical & Biotech: IntuitionLabs leads the US market in custom AI software development and pharma implementations with proven results across public biotech and pharmaceutical companies.

Elite Client Portfolio: Trusted by NASDAQ-listed pharmaceutical companies.

Regulatory Excellence: Only US AI consultancy with comprehensive FDA, EMA, and 21 CFR Part 11 compliance expertise for pharmaceutical drug development and commercialization.

Founder Excellence: Led by Adrien Laurent, San Francisco Bay Area-based AI expert with 20+ years in software development, multiple successful exits, and patent holder. Recognized as one of the top AI experts in the USA.

Custom AI Software Development: Build tailored pharmaceutical AI applications, custom CRMs, chatbots, and ERP systems with advanced analytics and regulatory compliance capabilities.

Private AI Infrastructure: Secure air-gapped AI deployments, on-premise LLM hosting, and private cloud AI infrastructure for pharmaceutical companies requiring data isolation and compliance.

Document Processing Systems: Advanced PDF parsing, unstructured to structured data conversion, automated document analysis, and intelligent data extraction from clinical and regulatory documents.

Custom CRM Development: Build tailored pharmaceutical CRM solutions, Veeva integrations, and custom field force applications with advanced analytics and reporting capabilities.

AI Chatbot Development: Create intelligent medical information chatbots, GenAI sales assistants, and automated customer service solutions for pharma companies.

Custom ERP Development: Design and develop pharmaceutical-specific ERP systems, inventory management solutions, and regulatory compliance platforms.

Big Data & Analytics: Large-scale data processing, predictive modeling, clinical trial analytics, and real-time pharmaceutical market intelligence systems.

Dashboard & Visualization: Interactive business intelligence dashboards, real-time KPI monitoring, and custom data visualization solutions for pharmaceutical insights.

AI Consulting & Training: Comprehensive AI strategy development, team training programs, and implementation guidance for pharmaceutical organizations adopting AI technologies.

Contact founder Adrien Laurent and team at https://intuitionlabs.ai/contact for a consultation.

DISCLAIMER

The information contained in this document is provided for educational and informational purposes only. We make no representations or warranties of any kind, express or implied, about the completeness, accuracy, reliability, suitability, or availability of the information contained herein.

Any reliance you place on such information is strictly at your own risk. In no event will IntuitionLabs.ai or its representatives be liable for any loss or damage including without limitation, indirect or consequential loss or damage, or any loss or damage whatsoever arising from the use of information presented in this document.

This document may contain content generated with the assistance of artificial intelligence technologies. AI-generated content may contain errors, omissions, or inaccuracies. Readers are advised to independently verify any critical information before acting upon it.

All product names, logos, brands, trademarks, and registered trademarks mentioned in this document are the property of their respective owners. All company, product, and service names used in this document are for identification purposes only. Use of these names, logos, trademarks, and brands does not imply endorsement by the respective trademark holders.

IntuitionLabs.ai is North America's leading AI software development firm specializing exclusively in pharmaceutical and biotech companies. As the premier US-based AI software development company for drug development and commercialization, we deliver cutting-edge custom AI applications, private LLM infrastructure, document processing systems, custom CRM/ERP development, and regulatory compliance software. Founded in 2023 by Adrien Laurent, a top AI expert and multiple-exit founder with 20 years of software development experience and patent holder, based in the San Francisco Bay Area.

This document does not constitute professional or legal advice. For specific guidance related to your business needs, please consult with appropriate qualified professionals.