Paper SA15

Harnessing AI to Streamline ClinicalTrials.gov PRS Submissions

Michele Andrush, Certara, USA

Anaya Rehman, Certara, Canada

ABSTRACT

The Protocol Registration and Results System (PRS) is a cornerstone of clinical trial transparency, yet manual submission processes remain a significant bottleneck for drug developers. This paper explores a transformative approach to ClinicalTrials.gov compliance by splitting the workflow into two AI-augmented pillars: Data Extraction and Structured Data Entry. By leveraging specialized Large Language Models (LLMs) and Human-in-the-Loop validation, study sponsors can reduce administrative burdens by up to 50% while improving data accuracy. We present a roadmap for transitioning from resource-intensive manual transcription to an automated, schema-compliant submission model that ensures regulatory readiness and end-to-end transparency.

INTRODUCTION

The Protocol Registration and Results System (PRS) serves as the primary gateway for clinical trial transparency, mandated by federal law under Section 801 of the Food and Drug Administration Amendments Act (FDAAA 801). This regulatory framework requires study sponsors or designated responsible parties to register applicable clinical trials (ACTs) no later than 21 days after the enrollment of the first participant [1]. Furthermore, summary results information must be submitted to ClinicalTrials.gov within 12 months of the primary completion date, defined as the date of final data collection for the pre-specified primary outcome measure [1]. These mandates, enforced by potential civil monetary penalties and the withholding of federal funding, ensure that critical trial data is accessible to patients, healthcare providers, and the global scientific community.

However, the current landscape of clinical disclosure is characterized by significant administrative complexity and data quality obstacles. Navigating the PRS interface manually requires teams to extract highly specific data points from voluminous source

documents, such as Clinical Study Protocols (CSPs) and Statistical Analysis Plans (SAPs), and transcribe them into a structured format. Research indicates that these manual processes are frequently plagued by incompleteness, with critical fields such as outcome measures, study design, and contact information often missing or underspecified [2]. Inconsistencies between the protocol and the registry record, as well as the usage of non-standardized free text instead of controlled ontologies like MeSH, further impair the reusability and credibility of the data [3].

THE STATUS QUO: CHALLENGES OF MANUAL PRS SUBMISSIONS

The traditional approach to ClinicalTrials.gov submissions is inherently resource intensive. Clinical disclosure experts must painstakingly review lengthy CSPs to identify required registration elements. This manual extraction phase is not only slow but represents a high administrative burden that may lead to submission delays.

Transcription errors, while often minor, can trigger rejection warnings from the National Library of Medicine (NLM), requiring multiple rounds of corrections and re-submissions. More critically, the manual process risks introducing inconsistencies between the technical language of the protocol and the summary information provided in the registry. These discrepancies can undermine the transparency goals of the registry and complicate the work of regulatory reviewers, researchers, sponsors and healthcare professionals. The manual method often struggles with data formatting inconsistencies when dealing with multiple vendors, making it difficult to achieve a single source of truth for all clinical data.

THE OBJECTIVE: AI-DRIVEN EFFICIENCY

Nearly 73% of life science companies still struggle to adopt appropriate AI technologies because traditional models are often trained on broad, non-specific datasets [4]. By leveraging AI tailored for drug development, companies can reduce the time spent collecting data and improve the overall research accuracy [5]. These platforms are often ISO27001 certified and can be deployed within a company's internal infrastructure, ensuring that sensitive clinical trial data remains secure and compliant with global privacy regulations. The shift toward an AI-driven solution is not about replacing human expertise but about harnessing AI-Driven Efficiency to manage the mechanical aspects of registration. The primary goal is to reduce manual effort while simultaneously increasing data accuracy. This approach aligns with the FDA's increasing support for automated and cloud-based solutions to streamline the clinical trial process [6].

The architecture of this modern workflow is split into two distinct phases:

1. **Intelligent Data Extraction**: Using AI to "read" source documents.

2. **Structured Data Entry**: Mapping that information to the required PRS fields.

Pillar One: Intelligent Data Extraction

In the first pillar, the AI system ingests the CSP and the SAP. Unlike traditional keyword searches, modern AI utilizes unique, targeted prompts to extract specific data points from the text. This allows the system to parse complex clinical language and identify nuances in study design or eligibility criteria.

A significant hurdle in this phase has historically been the format of the data. Many protocols are stored in non-searchable PDF files, creating an "analog" barrier. AI-augmented workflows can digitize this protocol data, ensuring that every relevant sentence is available for processing. Crucially, by constraining the AI to the source material, teams can ensure the system extracts data without "hallucinating" or inventing new content. This automated collection and analysis of unstructured clinical trial data enables faster, better trial design and execution.

Pillar Two: Structured Data Entry and Validation

Once the data is extracted, the AI maps the text directly to ClinicalTrials.gov fields, such as Study Design, Outcome Measures, and Arms/Interventions. This mapping ensures that there is schema compliance, meaning the data is formatted exactly as required by the PRS interface.

Built-in validation checks are then performed to identify any formatting issues. This proactive validation gives teams complete confidence that their data conforms with regulatory requirements before it is submitted, significantly reducing transcription warnings and rejection delays.

THE "HUMAN-IN-THE-LOOP" NECESSITY

A vital takeaway from early implementation is that *automation is not autopilot*. Human expertise remains essential as a validation layer. AI outputs may occasionally contain minor inconsistencies that require prompt refinement or manual correction. Furthermore, AI may still struggle with highly complex trial designs that require a deep, nuanced understanding of clinical methodology.

The optimal workflow is transactional processing, where the AI drafts the entry, but the clinical disclosure team provides the final oversight and review. This Human-in-the-Loop validation ensures that appropriate control, oversight, and compliance are maintained throughout the lifecycle of the submission. Expert oversight is also crucial when using AI for

related tasks like data anonymization and redaction to protect sensitive patient information.

MEASURING THE GAINS: EFFICIENCY, QUALITY, AND FUTURE-PROOFING

The transition from manual processes to an AI-augmented workflow yields measurable improvements across several critical performance metrics. Most notably, our initial testing showed a dramatic reduction in the hours traditionally dedicated to manual extraction and data entry, with first-draft creation timelines slashed by as much as 50%. This shift doesn't just save time; it fundamentally enhances the accuracy and quality of the output by minimizing transcription errors and ensuring tighter alignment with the source protocol.

Furthermore, high-quality metadata management directly translates to superior trial data, which bolsters regulatory readiness. This allows teams to meet unforgiving submission windows and respond to NLM queries with greater agility, ensuring filings remain on track. Ultimately, by stripping away the necessity for repetitive, labor-intensive tasks, automation drives significant cost reductions across the entire study lifecycle.

IMPLEMENTATION ROADMAP: A STRATEGIC PATH FORWARD

For an organization looking to implement an AI-augmented PRS process, the journey begins at the start with a structured, incremental strategy rather than an overnight overhaul. This initial phase requires a thorough audit of existing data silos to identify exactly where protocol data resides and to ensure it is properly formatted for AI ingestion. Once that data is accessible, the focus shifts to mapping protocol data, where clinical documents are directly aligned with the specific data points required by the PRS.

To ensure long-term success, starting small with a pilot program is highly recommended. By applying the AI-augmented workflow to a single study, a clear ROI can be demonstrated and internal protocols refined before a full-scale rollout. During this pilot phase, incorporating standardized assets is essential to ensure they can be reused efficiently across all future studies. Once the concept is proven, the scope can be expanded from initial registrations to results posting, ensuring compliance across global registries.

CONCLUSION

By shifting from a manual process to an AI-supported one, teams can future-proof their PRS workflows, ensuring they are prepared for the increasing transparency demands of the clinical research industry. The goal is to evolve the role of disclosure experts from data entry clerks to strategic gatekeepers of clinical integrity.

REFERENCES

1. FDAAA 801 and the Final Rule | ClinicalTrials.gov
2. Viergever, R. F., et al. (2014). The Quality of Registration of Clinical Trials: Still a Problem. *PLoS ONE*, *9*(1), e84727. https://pmc.ncbi.nlm.nih.gov/articles/PMC3044717/
3. Zarin, D. A., et al. (2007). Issues in the Registration of Clinical Trials. *JAMA*, *297*(19), 2112–2120. https://pubmed.ncbi.nlm.nih.gov/17507347/
4. Certara Blog, 2025. Beyond ChatGPT for Life Sciences: AI in Drug Discovery & Development
5. Bechelli, M. L., et al. (2025). ImpACT Project: Improving Access to Clinical Trials in Victoria, an Artificial Intelligence–Based Approach. *JCO Clinical Cancer Informatics*. https://pmc.ncbi.nlm.nih.gov/articles/PMC11732263/
6. U.S. Food and Drug Administration. (2023). *Using Artificial Intelligence & Machine Learning in the Development of Drug and Biological Products*. https://www.fda.gov/media/167936/download

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

- **Michele Andrush:** michele.andrush@certara.com
- **Anaya Rehman:** anaya.rehman@certara.com