ARTIFICIAL INTELLIGENCE-ASSISTED PROCESSING OF CLINICAL TRIAL SAFETY REPORTS: A PILOT IMPLEMENTATION STUDY

A. Williams, B. Neter, K. Lauzier, S. Han
Northwell Health Cancer Institute

BACKGROUND

Clinical trial safety reports require labor-intensive manual processing for data extraction and Principal Investigator review. Baseline assessment demonstrates 1.3 minutes per report processing time, which at institutional volumes of 500-1,000 monthly reports creates significant operational burden.

Per FDA regulations (21 CFR 312.60) investigators should review all IND safety reports received from sponsors as a part of the investigator's responsibility to protect the rights, safety, and welfare of trial participants.

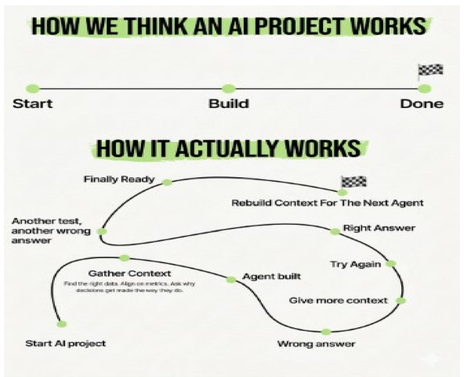
Safety report processing also generates substantial institutional revenue through per-report billing ($50-150/report, $50,000-150,000 annually), but manual workflows constrain capacity and delay compliance timelines. Artificial intelligence platforms offer potential for automation, but optimal implementation strategies in regulated research environments remain undefined.

OBJECTIVES

To develop and validate a standardized AI-driven methodology for processing clinical trial safety reports that eliminates the need for manual data extraction.

METHODS

This pilot study evaluated AI-assisted processing of Council for International Organizations of Medical Sciences (CIOMS) safety reports using the Northwell AI Hub platform (Gemini primary model version 2.5 Flash and 2.5 Pro). Two researchers, each independently processed safety reports manually versus AI assisted to generate cumulative monthly summaries of external reports, comparing processing times and accuracy. Prompt refinement was conducted iteratively with cross-validation between researchers. Variables assessed included report format variations (with/without summary cover pages), batch upload capacity, and output accuracy.



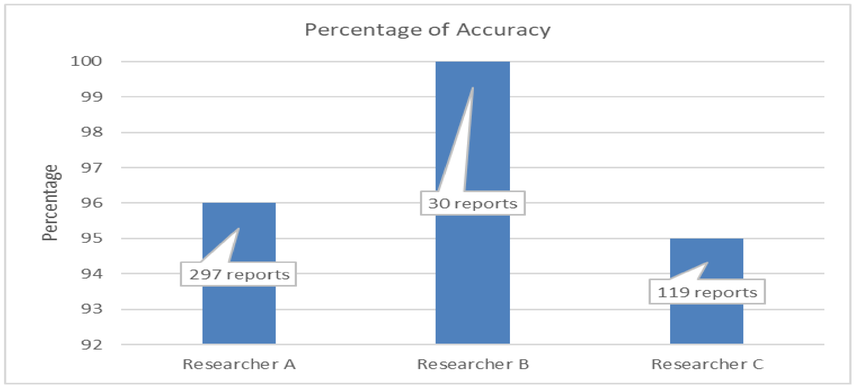
RESULTS

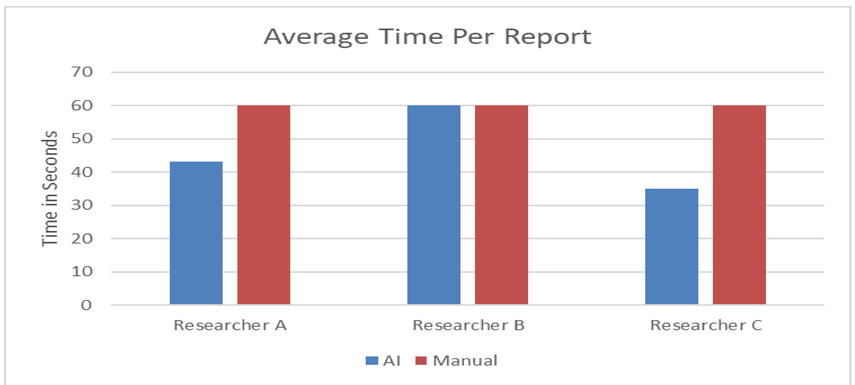
Researchers 1 & 2	AI-Assisted
Reports Processed: 20	Reports Processed: 20
Time Taken: 20 minutes (1 report / minute)	Time Taken: 3 minutes (automated summary)
100% Accuracy	100% Accuracy

Manual data extraction for site processing required 23-25 minutes for 20 reports at a rate of approximately 1 report/minute for both Researcher 1 and Researcher 2. Initial AI prompt optimization occurred over a 2-month period with 15 prompts generated to achieve consistent accuracy for one study-specific format, a Phase 1 trial with a summary cover page. Gemini 2.5 Pro AI model generated a markdown summary table in 3 minutes for 20 reports with 100% accuracy.

Previously the Gemini 2.5 Flash AI model was able to generate a markdown summary table in the same amount of time but required additional Researcher time for training of the model and prompt specificity to achieve the same level of accuracy. Further testing of CIOMS safety reports without summary cover pages utilizing the same prompt in the Gemini 2.5 Pro AI model demonstrated potential generalizability to multiple clinical trials. Depending on the model utilized, system limitations were identified at >40 report batch uploads, with the AI model producing either an error message or taking longer to produce results with 80 reports. Cross-platform validation across Gemini, ChatGPT, and Claude AI models is ongoing.

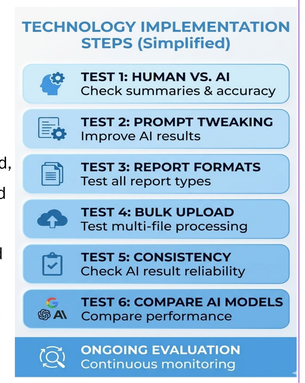
The pilot project was expanded, engaging three regulatory team members to evaluate AI prompts for CIOMS safety reports, encompassing variations with and without summary cover pages. A total of 446 reports were processed via AI, yielding an average accuracy of 97%. Notably, two of the three researchers demonstrated decreased processing time for AI summary generation and subsequent manual accuracy review, improving upon the manual average of one report per minute.





CONCLUSION

This pilot implementation demonstrates that AI-assisted processing of clinical trial safety reports can substantially reduce manual administrative burden while enabling scalable operations without increasing staffing requirements. By eliminating routine data extraction tasks, the AI-enabled workflow supports rapid, synthesis of safety information and allows highly skilled research personnel to be redeployed toward higher-value regulatory and scientific activities. These findings provide a compelling proof of concept for the responsible integration of artificial intelligence into clinical research operations and highlight its potential to modernize safety oversight processes across cancer centers.



LIMITATIONS

Format dependence: Performance varied by CIOMS report structure; absence of summary cover pages required additional prompt refinement, limiting immediate generalizability across sponsor templates.

System capacity constraints: Batch processing performance degraded or failed beyond ~40 reports, restricting throughput during high-volume periods.

Safety Report document quality: scanned documents with poor image quality impacted AI model's ability to read and extract data

Model variability: Accuracy, prompt sensitivity, and output consistency differed across AI models and platforms, limiting interchangeability and requiring model-specific optimization.

Ongoing human oversight: Expert review remained necessary to verify outputs and ensure compliance with investigator safety review requirements.

IMPLICATIONS

AI-assisted external safety report processing demonstrates promise for improving operational efficiency and reducing administrative regulatory burden within cancer center research environments, though current performance remains dependent on format-specific prompt engineering. Ongoing challenges related to external safety report template heterogeneity and platform capacity highlight the need for standardized approaches. Future work will focus on developing universal prompting frameworks and conducting structured, comparative platform evaluations to support scalable and responsible adoption across regulated clinical research operations.