Context-aware Generative AI for Evidence Generation, Analog-driven Enrollment Forecasting, and Operational Risk Management in Oncology Trials

Ashwin Kumar Rai[1], Victoria Ikoro[2], Devika Bhandary[3], Andre Ng[3], Marielle Bassel[4]

[1]Thermo Fisher Scientific, Waltham, MA, USA; [2]Thermo Fisher Scientific, Kitchener, ON, Canada; [3]Thermo Fisher Scientific, London, UK; [4]Thermo Fisher Scientific, Montreal, QC, Canada

MSR60

Background

- Clinical trial feasibility planning depends on identifying historical trials with similar design and patient populations ("analogs").[1]
- Traditional analog selection is often manual, subjective, and difficult to scale, which can reduce reproducibility and consistency in feasibility assessments.[2]
- Public resources such as ClinicalTrials.gov contain rich trial information, but much of it is unstructured and not readily comparable across studies.[1]
- Advances in natural language processing and machine learning enable transformation of unstructured clinical trial text into standardized, structured features for systematic comparison.[3,4]
- Recent large language model approaches further demonstrate the value of artificial intelligence (AI) for extracting and matching clinical trial information from complex free text.[5]
- There remains a need for context-aware, reproducible frameworks that translate trial similarity into quantitative enrollment benchmarks and operational insights.[2,6]

Objectives

- Develop a context-aware GenAI workflow to identify oncology trials with matched study designs
- Extract enrollment performance and operational difficulty signals
- Generate explainable recommendations for enrollment rates in planned trials
- Establish a scalable pipeline to support future machine learning applications

Methods

Data and Study Selection

- Completed interventional oncology trials were identified from a structured clinical trial dataset. Key study attributes included trial phase, design, eligibility criteria, enrollment, and geographic footprint. Enrollment performance was quantified as patients per site per month (PPSM).

GenAI-based Trial Characterization

- A context-aware generative AI model was used to extract standardized features from unstructured trial text, including study title, conditions, interventions, and summaries. Each trial was represented as a structured "trial fingerprint," capturing key clinical and operational attributes such as indication, stage/line, biomarker requirements, comparator class, eligibility complexity, trial burden, and geographic footprint (**Table 1**).

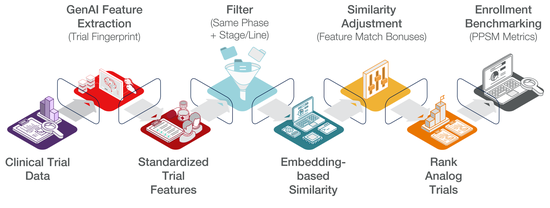
Table 1. Trial Features Used for Similarity Assessment

Feature	Description
Indication	Extracted disease category from trial text
Stage/Line	Disease stage (early, metastatic, etc.)
Phase	Clinical trial phase
Biomarker Required	Whether enrollment requires biomarker
Comparator Class	Trial design comparator type
Trial Burden	Intensity of procedures and visits
Eligibility Complexity	Restrictiveness of inclusion criteria
Geographic Footprint	Distribution of trial sites

Similarity and Analog Identification

- Trial fingerprints were converted into vector representations and compared using embedding-based similarity. Candidate analog trials were restricted to those with matching **trial phase and stage/line**, ensuring clinical comparability (**Figure 1 and Figure 3**).
- Similarity scores were further refined using rule-based adjustments for exact matches on key attributes, and top-ranked analog trials were selected for each target study.

Figure 1. Analog Identification and Enrollment Benchmarking Framework



Enrollment Benchmarking

- Enrollment rates from selected analog trials were summarized using median, interquartile range, minimum and maximum values, and similarity-weighted mean PPSM. These metrics were used to generate **data-driven enrollment expectations**.

Methods (continued)

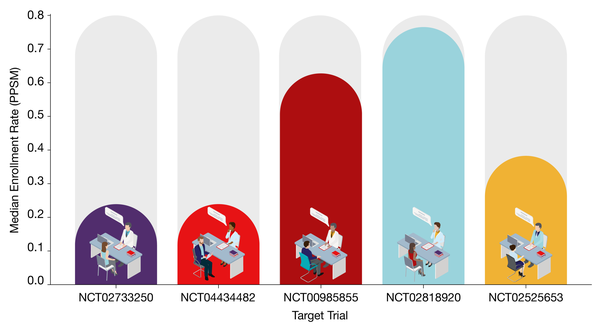
Explainability and Study Design

- The framework generated qualitative insights describing factors expected to influence enrollment performance, including biomarker requirements, eligibility complexity, comparator design, and trial burden.
- Five completed oncology trials were treated as target studies, each matched to up to 10 analog trials.

Results

- Across five target oncology trials, the framework identified clinically relevant analog cohorts and generated enrollment benchmarks based on analog trial performance. Each target trial was matched to 4 to 10 analog trials, with analogs aligned on key design features including phase, stage/line, and indication.
- Median enrollment rates ranged from 0.24 to 0.77 PPSM, while similarity-weighted mean enrollment rates ranged from 0.34 to 0.75 PPSM (**Figure 2**). Across targets, enrollment performance varied substantially, indicating that even clinically similar trials can have meaningfully different recruitment patterns.

Figure 2. Median Enrollment Rates (PPSM) Across Target Trials



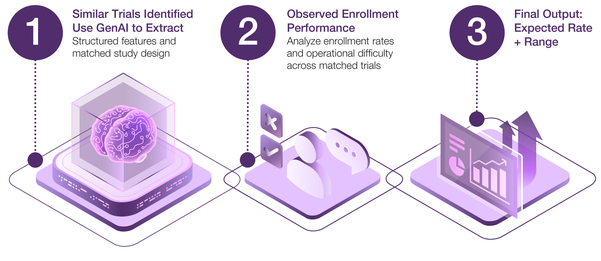
Abbreviation: PPSM = patients per site per month

- Enrollment benchmarks derived from analog trials showed meaningful variation across clinical settings.
- Median PPSM ranged from 0.24 to 0.77; weighted mean PPSM ranged from 0.34 to 0.75.
- The overall enrollment range across analogs is 0.03 to 2.93 PPSM.
- These findings indicate that even among clinically similar trials, enrollment performance can vary substantially, highlighting the importance of capturing both central tendency and variability.

Cross-target Comparison of Enrollment Performance

- Clear differences in expected enrollment were observed across trial settings:
  - Neoadjuvant non-small cell lung cancer (NSCLC) (NCT02818920) demonstrated the highest enrollment performance, with a median PPSM of 0.77 and a weighted mean of 0.75, suggesting favorable recruitment conditions in earlier-stage disease settings.
  - Locally advanced lung cancer (NCT00985855) showed moderate enrollment performance, with a median PPSM of 0.63, reflecting a balance between clinical complexity and patient availability.
  - Metastatic NSCLC and small cell lung cancer targets (NCT02733250 and NCT04434482) demonstrated the lowest enrollment rates, both with a median PPSM of 0.24, suggesting greater recruitment challenges in advanced disease settings.
- These results highlight that trial context (stage/line and disease setting) is a key determinant of enrollment performance.

Figure 3. How Analog Trials Can Inform Enrollment Expectations



Results (continued)

Variability in Analog Trial Performance

- Despite high similarity between target and analog trials, substantial variability in enrollment performance was observed within analog cohorts. The most notable example was NCT02525653 (metastatic lung cancer); median PPSM was 0.38, and the interquartile range was 0.09 to 0.99 (**Table 3**).
- The full range was 0.03 to 2.93, and the weighted mean was 0.69. This represents nearly a 100-fold difference between the lowest and highest performing analog trials, even within a highly similar cohort.
- This finding reinforces that single-trial comparisons are insufficient. Distribution-based benchmarking is necessary for realistic planning

Role of Similarity Weighting

- Similarity-weighted mean PPSM values were generally higher than median PPSM across targets, indicating that **closer-matching analog trials tend to exhibit stronger enrollment performance**.
- For example:
  - NCT02525653: median **0.38** versus weighted mean **0.69**
  - NCT02733250: median **0.24** versus weighted mean **0.34**
- This demonstrates that similarity weighting helps prioritize **more relevant analog evidence**, improving the interpretability of enrollment estimates.

Table 3. Summary of Enrollment Performance (PPSM) Across Target Trials Based on Matched Analog Cohorts

Target Trial	Indication	Stage/Line	Phase	N Analogs	Median PPSM	Q1–Q3 PPSM	Min–Max PPSM	Weighted Mean PPSM	Key Insight
NCT02733250	NSCLC	metastatic	Phase 1/2	8	0.24	0.16–0.54	0.06–0.76	0.34	Stable, moderate variability
NCT04434482	SCLC	metastatic	Phase 1/2	8	0.24	0.17–0.54	0.13–0.76	0.35	Higher procedure burden
NCT00985855	Lung (other)	locally advanced	Phase 2	5	0.63	0.18–0.90	0.15–1.07	0.59	Comparator + burden effects
NCT02818920	NSCLC	neoadjuvant	Phase 2	4	0.77	0.54–0.97	0.36–1.10	0.75	Favorable recruitment setting
NCT02525653	Lung (other)	metastatic	Phase 2	10	0.38	0.09–0.99	0.03–2.93	0.69	High variability

Abbreviations: NSCLC = non-small cell lung cancer; PPSM = patients per site per month; SCLC = small-cell lung cancer

Operational Insights and Explainability

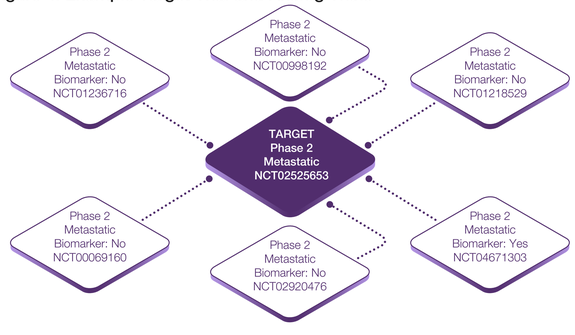
- In addition to quantitative benchmarks, the framework generated qualitative insights explaining observed enrollment patterns (**Table 4**):
  - High procedure burden and restrictive eligibility criteria were associated with reduced enrollment or increased variability.
  - Comparator design (e.g., active control or standard-of-care arms) influenced patient willingness to participate.
  - Randomized designs introduced additional recruitment challenges in certain settings.
  - Targets without major complexity signals demonstrated more stable enrollment performance, suggesting that operational simplicity contributes to improved recruitment outcomes.

Table 4. Factors Associated With Enrollment Performance

Lower Enrollment	**Observed in:** Metastatic NSCLC and SCLC **Key factors** • High procedure/visit burden • Complex eligibility criteria • Comparator design (active control) • Randomized study design
Moderate Enrollment	**Observed in:** Locally advanced lung cancer **Key factors** • Balanced clinical complexity • Moderate operational burden • Mixed enrollment constraints
Higher Enrollment	**Observed in:** Neoadjuvant NSCLC **Key factors** • Earlier-stage disease setting • Lower operational complexity • Fewer enrollment constraints

Abbreviations: NSCLC = non-small cell lung cancer; SCLC = small-cell lung cancer

Figure 4. Example Target Trial and Analog Trials



- Analog-based benchmarking reveals that even clinically similar trials can exhibit substantial variability in enrollment performance (**Figure 4**).
- By combining similarity-based trial selection with distribution-based metrics, this approach provides both an expected enrollment rate and a realistic range of outcomes.
- This enables more transparent, data-driven feasibility planning compared with traditional single-trial comparisons.

Strengths and Limitations

- This study demonstrates a context-aware framework that integrates generative AI with structured trial data to identify clinically relevant analog trials and generate enrollment benchmarks. The approach enables transformation of unstructured trial text into standardized, interpretable features, supporting consistent comparison across studies. By combining similarity-based ranking with distribution-based enrollment metrics, the framework provides both realistic expectations and insight into variability, while maintaining transparency through explainable feature contributions.
- However, several limitations should be considered. The demonstration is based on a limited number of target trials and may not capture the full range of variability across oncology studies. The accuracy of extracted features depends on the completeness and quality of source trial descriptions, and similarity adjustments rely on predefined rules rather than learned optimization. Additionally, substantial variability in enrollment performance persists even among highly similar trials, reflecting unmeasured operational and real-world factors not fully captured by the current framework.

Conclusions

- This study demonstrates that a context-aware GenAI framework can systematically identify clinically relevant analog trials and generate data-driven enrollment benchmarks for clinical trial planning.
- By combining similarity-based analog selection with distribution-based performance metrics, the approach provides both an expected enrollment rate and a realistic range of outcomes, addressing limitations of single-trial comparisons.
- The integration of quantitative outputs with qualitative explanations enhances interpretability and supports more informed feasibility decision-making.
- With further scaling, this framework has the potential to enable large-scale evidence generation and machine learning–based prediction of clinical trial performance.

References

1. ClinicalTrials.gov. U.S. National Library of Medicine. Accessed April 13, 2026. https://clinicaltrials.gov
2. Carlisle B, et al. *Clin Trials*. 2015;12(1):77-83.
3. Idnay B, et al. *J Am Med Inform Assoc*. 2022;29(1):197-206.
4. Rajkomar A, et al. *N Engl J Med*. 2019;380(14):1347-1358.
5. Jin Q, et al. *Nat Commun*. 2024;15:9074.
6. Bieganek C, et al. *PLoS One*. 2022;17(2):e0263193.

Disclosures

Funding is provided by Thermo Fisher Scientific. **AK, VI, DB,** and **AN** are employees of PPD™ Evidera™ Real-World Data & Scientific Solutions, Thermo Fisher Scientific. **MB** is an employee of PPD Prospective Real-World Studies, Thermo Fisher Scientific.

Acknowledgments

Editorial and graphic design support were provided by Amara Tiebout, Richard Leason, and Kawthar Nakayima of Evidera, a business unit of PPD, part of Thermo Fisher Scientific.

© 2026 Thermo Fisher Scientific Inc. All rights reserved. All trademarks are the property of Thermo Fisher Scientific and its subsidiaries unless otherwise specified. This information is not intended to encourage use of these products in any manner that might infringe the intellectual property rights of others.