

CLINICALTRIALSHUB: Bridging Registries and Literature for Comprehensive Clinical Trial Access

Jiwoo Park Ruoqi Liu Avani Jagdale Andrew Srisuwananukorn
 Jing Zhao Lang Li Ping Zhang Sachin Kumar

The Ohio State University, Columbus OH

park.3620@osu.edu

Abstract

We present CLINICALTRIALSHUB, an interactive search-focused platform that consolidates all data from ClinicalTrials.gov and augments it by automatically extracting and structuring trial-relevant information from PubMed research articles. Our system effectively increases access to structured clinical trial data by 83.8% compared to relying on ClinicalTrials.gov alone, with potential to make access easier for patients, clinicians, researchers, and policymakers, advancing evidence-based medicine. CLINICALTRIALSHUB uses large language models such as GPT-5.1 and Gemini-3-Pro to enhance accessibility. The platform automatically parses full-text research articles to extract structured trial information, translates user queries into structured database searches, and provides an attributed question-answering system that generates evidence-grounded answers linked to specific source sentences. We demonstrate its utility through (1) a user study involving clinicians, clinical researchers, and PhD students of pharmaceutical sciences and nursing, and (2) a systematic automatic evaluation of its information extraction and question answering capabilities.¹

1 Introduction

Access to clinical trial information is essential for patients seeking new treatments and for clinicians, researchers, and policymakers working to advance medical care. Containing over 500K registrations, [ClinicalTrials.gov](#) (CTG) is the primary resource many of these groups use to search and identify ongoing and completed trials. All trials in CTG

¹The demonstration is hosted at: <https://cth-frontend-103266204202.us-central1.run.app/>. A demonstration video is available at: <https://www.youtube.com/watch?v=uCPxyw7Abh0>. The source code to run the demo locally is available at: <https://github.com/jiwoo-jus/clinical-trials-hub>. Evaluation code for comparing model performance is available at: <https://github.com/jiwoo-jus/clinical-trials-hub-evaluation>.

Database	Filter/Status	Count
CTG	All Studies	543,172
CTG	Completed w/ Results	60,449
PubMed	Max Sensitivity ^a	962,774
PubMed	Max Specificity ^a	138,279
PubMed	Max Sensitivity + CTG linked ^b	63,928
PubMed	Max Specificity + CTG linked ^b	35,671

Table 1: Clinical trial data coverage comparison. ^aOpen-access trials identified using PubMed Clinical Study Categories filters from Haynes et al. ([Wilczynski et al., 2011](#)), with PMC Open Access restriction. ^bSubset of trials already registered on ClinicalTrials.gov.

are available in a structured format allowing the users to easily navigate, filter, or download them. However, many trials remain unregistered or are reported only in publications, particularly those conducted outside the US. [PubMed](#), which indexes over 35 million biomedical papers and abstracts, often includes results from these unregistered trials. As shown in Table 1, a substantial number of trials accessible from PubMed are not registered on CTG. The difference in coverage reflects the lack of integration between trial registries and published literature. Despite both being critical resources, PubMed and CTG exist as isolated silos with incompatible formats. PubMed, containing free-text articles, can be especially difficult for non-researchers to parse or filter, requiring substantial manual effort that could take weeks or months.

To address this gap, we present CLINICALTRIALSHUB, a unified user-centered trial search platform that combines trial registry data from CTG and published literature from PubMed into a single interface (§2). Our system makes three main contributions. First, our interface accepts natural language queries and displays unified search results combining trials from both sources in a single ranked list, merging overlapping trials (§3). Sec-

ond, using large language models (LLMs), we extract **structured information** from free text PubMed articles making up to 899,846 previously unregistered clinical trials easily searchable and filterable (§4). Lastly, we build an attributed **question answering** feature that allows the users to ask questions over individual trials (§5); answers are generated with attribution provided in the trial text.

Our system showcases practical use cases spanning diverse medical contexts, from evidence curation and systematic reviews to dosing protocols and treatment evaluation. We validate the system’s utility through comprehensive evaluations comparing frontier LLMs (Gemini-3-Pro ([Gemini Team, Google, 2025](#)), GPT-5.1 ([OpenAI, 2025](#)), Claude-4.5-Sonnet ([Anthropic, 2025](#))) to select optimal models for each feature. We conduct quantitative benchmarking to assess information extraction accuracy using a curated benchmark and evaluate question-answering quality using the FACTS grounding benchmark ([Jacovi et al., 2025](#)). Additionally, we perform a user study with seven medical professionals. The evaluations confirm the platform’s effectiveness across its core functionalities and showcase its utility over using PubMed or CTG alone.

2 CLINICALTRIALSHUB UI Design

This section describes the user interface of our system. The platform consists of two primary pages: a main search page and individual trial detail pages.

Main Search Page. The main search interface (an overview is shown in [Figure 1](#); additional screenshots are available in [Appendix G](#)) offers a unified search experience. In **search bar panel**, users can enter natural language queries (top) or use structured input forms specifying condition, intervention, and other terms for PICO-based search (middle), or utilize expert-level PubMed/CTG query forms (bottom). The **search results** display a ranked list of trials from both CTG and PubMed sources, with merged entries when a trial appears in both databases. A **filtering sidebar** enables refinement by data source, study type, phase, design allocation, and other trial-specific categories. The **details sidebar** provides a quick preview of abstracts and metadata when clicking trial titles. CLINICALTRIALSHUB also provides two experimental features to enhance the search experience. Users can specify **inclusion and exclusion criteria** for early-stage screening; when viewing a trial’s

preview sidebar, the system provides a quick eligibility assessment based on the abstract and metadata. Additionally, an **AI-insight** feature generates five context-aware insights based on the current search results, supporting iterative Q&A by maintaining context from the user’s queries and recent interactions.

Individual Trial Detail Page. Upon selecting a trial, users access a comprehensive detail page (an overview is shown in [Figure 2](#); additional screenshots are provided in [Appendix H](#)) with three main components evaluated in our studies: **Structured Information** displays trial metadata in an easy-to-navigate format. This information is directly available for registered trials on CTG, while for PubMed-only items it is extracted from free-text articles; **Full Text** shows the entire publication text for PubMed entries. For CTG entries, referenced PMC articles are listed, and these can then be clicked and expanded to full text as well. **Interactive QA** allows users to ask questions about the trial, with answers generated by an LLM fed PMC articles or CTG structured data depending on the item, and generates evidence-based responses. When users click the evidence, it auto-scrolls and highlights upon clicking a citation for verification.

3 CLINICALTRIALSHUB Search

Our search system comprises three components: (1) query refinement that converts natural language queries to platform-specific structured queries, (2) multi-source search on the two platforms, (3) relevance reranking and deduplication. [Figure 3](#) illustrates the overall pipeline.

3.1 Query Refinement

The query refinement component translates user queries into structured search parameters. Our implementation handles both structured field inputs and natural language queries. When users provide individual field values (condition, intervention, other terms), we use them directly. For natural language queries, we use an LLM (GPT-5.1) that applies clinical terminology normalization to decompose the required fields. We design prompt templates to extract condition terms, intervention specifications, and auxiliary keywords. The prompt used to extract these meanings from user queries is shown in [Appendix I](#).

The extracted output maps directly to database-specific query construction, where PubMed

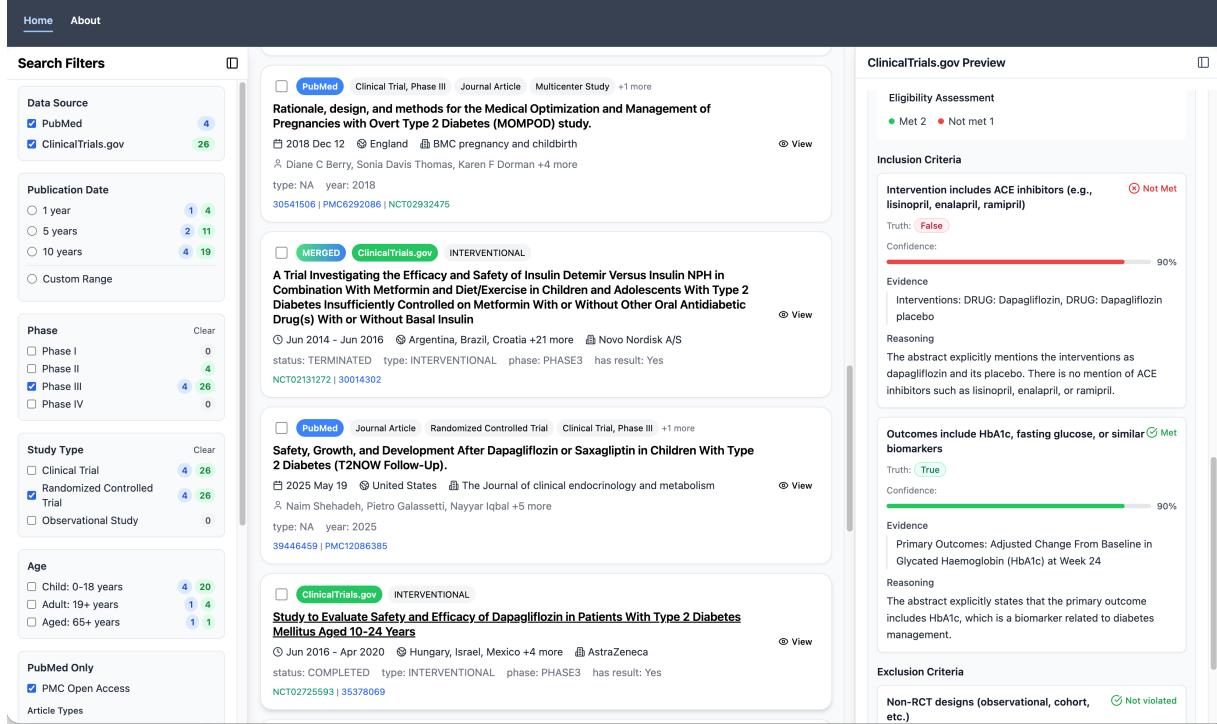


Figure 1: Main search interface with unified results, filters, and eligibility preview

searches combine the structured parameters with Boolean operators and predefined clinical trial detection patterns, while CTG requires the decomposed fields as separate API parameters. This approach maintains consistent query semantics across heterogeneous database interfaces while preserving the original search intent.

3.2 Multi-Source Search

Both CTG and PubMed offer their own search APIs. Give the refined query, we execute both APIs asynchronously with the following platform specific optimization strategies.

PubMed employs a two-stage retrieval process using NCBI's [E-utilities API](#).

- **E-SEARCH:** executes optimized Boolean queries against PubMed's index to extract relevant PMIDs filtering for journal restrictions, demographic constraints, and condition-specific query enhancement.
- **E-FETCH:** performs batch metadata retrieval for comprehensive bibliographic information, including structured abstracts, MeSH terms, chemical substances, grant information, and clinical trial cross-references.

CTG Integration utilizes a dual-source approach for comprehensive trial data access:

- **Primary API:** Retrieves basic trial information and NCT identifiers via the official ClinicalTrials.gov REST API with support for complex filtering, including study type, phase, enrollment criteria, and geographical constraints.
- **AACT Database:** Leverages the daily-updated PostgreSQL mirror of ClinicalTrials.gov for enhanced metadata extraction, detailed outcome measures, and complex relational queries not supported by the primary API.

3.3 Reranking and Result Integration

We apply BM25 scoring ([Robertson and Walker, 1994](#)) separately to each source, normalize scores, and add bonuses based on the similarity ranking within each source. We bidirectionally merge PubMed and CTG entries that reference each other, providing a relevance bonus to merged pairs. All entries are then sorted by final scores and displayed to users. Complete technical details including corpus construction, scoring formulas, and deduplication algorithms are provided in [Appendix B](#).

4 Information Extraction

All entries on CTG follow a structured format that allows users to navigate and filter across different trials as well as download them for analyses. For example, a patient looking for phase 3 trials for

ChatBot

Q: Did the trial include participants who were already using basal insulin?

A: Yes, the trial included participants who were already using insulin as part of their standard of care. The inclusion criteria specified that eligible participants could be on a stable dose of insulin, either alone or in combination with metformin. Furthermore, the study protocol noted that basal insulin could be up-titrated as a rescue measure, implying its potential presence in the treatment regimen.

Evidence:

- Inclusion criteria were HbA1c 6.5–11%, FPG ≤14.2 mmol/L (≤ 256 mg/dL) and a stable dose of either metformin (>1000 mg daily), insulin, or a combination of metformin (>1000 mg daily) plus insulin for a minimum of 8 weeks.
- Basal insulin was initiated or up-titrated as open-label rescue for participants meeting prespecified criteria for lack of glycemic control

Ask about PMC10851108

Structured Information

Protocol Section Results Section Document Section Derived Section

Participant Flow

Recruitment Details

Participants took part in the study at 42 study centres in 7 countries worldwide.

Arm/Group Information

ID	Title	Description
FG000	Dapagliflozin 10mg/ Dapagliflozin 10mg	Dapagliflozin (10 mg) tablet administered orally, once daily for the 24 week double-blinded treatment period. The participants then continued to receive Dapagliflozin (10 mg) once daily for a further 28 weeks in the open label long term-extension.
FG001	Placebo/ Dapagliflozin 10mg	Matching placebo tablet administered orally, once daily for the 24 weeks double-blinded treatment period. The participants then received Dapagliflozin (10 mg), orally, once daily for a further 28 weeks in the open label long-term extension.

Study Periods

Discrete stages of a clinical study during which numbers of participants at specific significant events or points of time are reported.

Arm/Group Title	Dapagliflozin 10mg/ Dapagliflozin 10mg	Placebo/ Dapagliflozin 10mg
Blinded Treatment Period		
Started	39	33
Received Treatment	39	33
Completed	34	27
Not Completed	5	6
Long Term Extension		
Started	33	27
Completed	32	24
Not Completed	1	3

Baseline Characteristics

A description of each baseline or demographic characteristic measured in the clinical study

Baseline Groups

ID	Title	Description
~~~~~	Dapagliflozin 10mg/	Dapagliflozin (10 mg) tablet administered orally, once daily for the 24 week double-blinded treatment

Figure 2: Trial detail page with structured information and evidence-grounded QA

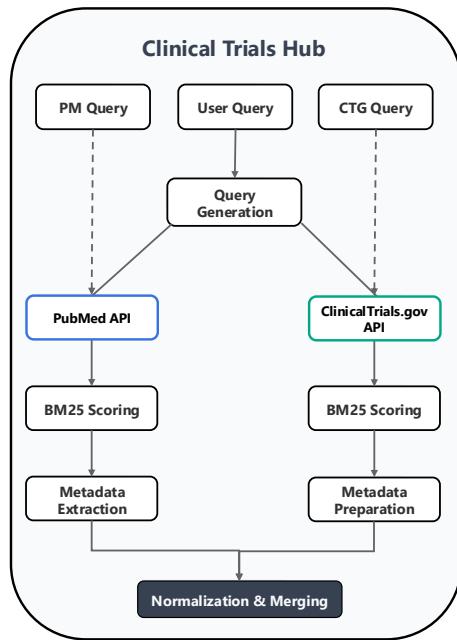


Figure 3: Search pipeline

their gender and age range can easily apply a filter and find relevant trials. On the other hand, PubMed articles are research papers in free text form which makes it harder to perform such tasks. Hence, we implement an information extraction pipeline that extracts all relevant trial-related data from papers in the same format as CTG, making indexing, filtering, browsing, and analysis through them easier. We employ an LLM for this purpose that takes as input (parts of) the paper and generates the structured output in a JSON format.

We perform information extraction for **211** fields in parallel by dividing the task into six modules for the protocol section, four modules for the results section, and one module for derived section, each with explicit field definitions and enumerated value constraints constructed based on CTG data schema² fed to the LLM (see fields list at [Table 14](#) in [Appendix C](#)). This approach minimizes both user wait times and hallucinations caused by the LLM processing overwhelming amounts of information. We describe them in detail in [Appendix C](#). The prompt templates are provided in [Appendix J](#)

#### 4.1 Evaluation

We evaluated three frontier LLMs (Gemini-3-Pro, GPT-5.1, and Claude-4.5-Sonnet) to measure struc-

²<https://beta-ut.clinicaltrials.gov/api/v2/studies/metadata>

tural completeness (key-level: whether the field exists) and semantic accuracy (value-level: whether the extracted value matches the ground truth).

**Dataset construction** We create an evaluation dataset comprising 100 PMC-CTG trial pairs. To do this, we queried PubMed for clinical trials ([subsection C.3](#)) published after 2021 (the current data schema’s modernization point) with PMC full-text availability and CTG status set to completed. We then check for a one-to-one mapping verified bidirectionally through the PMC and CTG databases. We focus on the **21** distinct field types in protocol section covering study design, arm group interventions, outcome measures, and eligibility criteria.

We excluded CTG fields appearing in fewer than 15% of records and fields primarily found in PMC articles but inconsistently updated on CTG, considering them unsuitable as reliable ground truth. Since results reporting has been required, limited enforcement has led to results information being more complete in PMC articles, while CTG entries may be missing or outdated. Conversely, CTG often lists extensive secondary outcomes and adverse events, whereas PMC articles emphasize primary outcomes. We plan to create a systematically validated benchmark dataset covering this results section as our future work. Model predictions and ground truth data are available [here](#).

For key-level evaluation, we flatten all fields from both model prediction and reference, preserving hierarchical relationships. For each field in the schema, we classify its extraction status as: *True Positive* (field exists in both prediction and reference), *False Positive* (field appears in prediction but not defined in the CTG data schema), *False Negative* (field exists in reference but missing from prediction), or *Extra Valid* (field appears in prediction and is valid according to the schema, but absent from this specific reference case, making it unevaluable for correctness).

For value-level evaluation, for fields with *True Positive* status, we assess extracted values using a three-stage approach. First, we perform exact string matching with normalization. For mismatches, we calculate semantic similarity using GPT-5.1 as judge with a 0.70 threshold (prompt shown in [subsection C.4](#)). We choose this threshold qualitatively by manually analyzing model outputs. For structured list fields, we create an  $m \times n$  similarity matrix where  $m$  is the number of reference elements and  $n$  is the number of predicted elements

to score all reference-prediction element pairs, then use the Hungarian algorithm ([Kuhn, 1955](#)) to optimize total similarity across matched pairs, allowing accurate per-element evaluation regardless of list order.

**Results** Table [Table 2](#) summarizes the overall performance. GPT-5.1 achieved the highest performance across all metrics, recording a Key-level F1 of 0.980 and a Value-level F1 of 0.890. Gemini-3-Pro exhibited competitive performance (Key F1 0.960, Value F1 0.870), whereas Claude-4.5-Sonnet showed a clear performance gap (Key F1 0.740, Value F1 0.660). However, we note that the differences between GPT-5.1 and Gemini-3-Pro are small across both key- and value-level metrics. Because we used GPT-5.1 as the semantic similarity judge, a small bias in its favor is possible. To mitigate this, we manually reviewed some of the disagreement cases and confirmed that the judged matches were generally valid. Given the overall closeness of model performance, our results suggest that either GPT-5.1 or Gemini-3-Pro could serve as a reasonable backbone model in practice. We select GPT-5.1 for our system primarily for cost-efficiency and practical deployment considerations rather than because of any substantial performance gap.

A more fine-grained view is presented in [Table 13](#), which reports value-level F1 scores for each individual field alongside the average number of Extra-Valid (EV) cases. Although EV cases cannot be directly evaluated due to the lack of a reference value, they nevertheless reveal meaningful patterns about the underlying data sources. Several descriptive and eligibility-related fields exhibit high EV frequencies, indicating that PMC articles often contain information that is absent in CTG. Importantly, these same fields also show high F1 scores when CTG does provide a reference value, suggesting that discrepancies arise primarily from CTG incompleteness rather than model hallucination. This pattern reinforces the complementary value of LLM-based extraction when dealing with fields that are inconsistently maintained or only partially populated in CTG.

## 5 Interactive Grounded QA

On the trial details page, we provide an LLM-powered chat interface that allows the user to ask question about individual trials with evidence-grounded responses. To select the optimal model

Level	Model	Precision	Recall	F1
Key	Gemini-3-Pro	<b>1.000</b>	0.930	0.960
	GPT-5.1	<b>1.000</b>	<b>0.970</b>	<b>0.980</b>
	Claude-4.5-Sonnet	0.900	0.690	0.740
Value	Gemini-3-Pro	0.830	0.910	0.870
	GPT-5.1	<b>0.840</b>	<b>0.960</b>	<b>0.890</b>
	Claude-4.5-Sonnet	0.740	0.670	0.660

Table 2: Information extraction performance across key-level and value-level metrics.

for this component, we evaluated three frontier LLMs (Gemini-3-Pro, GPT-5.1, Claude-4.5-Sonnet).

**Benchmark Selection** For this evaluation, we used the FACTS Grounding benchmark (Jacovi et al., 2025), which measures a model’s ability to generate responses factually grounded in provided context documents. The benchmark aligns directly with our requirements by incorporating substantial medical-domain coverage (29%), supporting long context documents up to 32K tokens with diverse query patterns, and explicitly constraining models to rely solely on the provided context.

**Evaluation Methodology** FACTS employs a two-phase evaluation conducted by three judge models(Gemini-1.5-Pro, GPT-4o, and Claude-3.5-Sonnet). **1. Instruction Following (Eligibility):** Judges assess whether responses adequately address the user request, assigning verdicts of *No issues*, *Minor issue(s)*, or *Major issue(s)*. A response is deemed ineligible (and assigned a score of 0) if all three judges classify it as having major issues, preventing it from advancing to the grounding evaluation phase. **2. Grounding:** Judges evaluate whether each response is fully grounded in the source document by classifying every sentence as *supported*, *unsupported*, *contradictory*, or *no_rad* (requires no factual grounding). A response receives a positive grounding verdict from a judge only if all sentences are either supported or *no_rad*. The factuality score for each response is the average of the three judges’ verdicts, and each model’s final score is the mean factuality score across all evaluated responses.

**Evaluation setup** We evaluated Gemini-3-Pro, GPT-5.1, and Claude-4.5-Sonnet on 236 medical-domain samples from FACTS, generating all responses but evaluating the top 100 shortest average response length to balance cost and significance following FACTS guideline.

Model \ Judge	Judge	Average	Gemini	GPT	Claude
Gemini-3-Pro	<b>0.897</b>	<b>0.920</b>	<b>0.830</b>	<b>0.930</b>	
GPT-5.1	0.680	0.670	0.610	0.760	
Claude-4.5-Sonnet	0.767	0.740	0.690	0.860	

Table 3: Evaluation scores by judge model. Each row shows a prediction model evaluated by three different judge models (Gemini-3-Pro, GPT-5.1, Claude-4.5-Sonnet) and their average score.

For grounding evaluation, We used the *JSON* prompt template from the available templates, which FACTS identified as optimal for Gemini-1.5-Pro and GPT-4o. This JSON template instructs the model to output structured JSON objects for each sentence, with classification labels, explicit rationales explaining each decision, and supporting evidence excerpts from the context document.

In contrast, the *implicit-span-level* prompt template generates unstructured natural language output, listing each sentence with a simple binary accurate/inaccurate label and concluding with a final verdict, without providing detailed rationales or supporting evidence. In FACTS, this approach had previously been preferred for Claude-3.5-Sonnet due to its lower complexity in structured output generation. However, Claude-4.5-Sonnet exhibited substantially improved JSON generation reliability, allowing us to use the more informative template format. (see Table 15)

**Results** Gemini-3-Pro achieved the highest grounding score (0.897), notably outperforming GPT-5.1 (0.680) and Claude-4.5-Sonnet (0.767), with consistently stronger performance across all three evaluators (see Table 3). We additionally recorded completion tokens, response time, and response length for all generated responses. As shown in Table 4, while Gemini-3-Pro required significantly higher completion tokens (1,318 avg. vs. ~220 for others) and generation time (14.33s vs. ~4s), it produced the most concise final answers (698 characters vs. ~950). This suggests that Gemini performs a deeper internal reasoning process to synthesize grounded, compact responses; this behavior is critical for trustworthy clinical QA. Consequently, we selected Gemini-3-Pro as our QA backbone.

## 6 User Study

To assess the practical utility of CLINICALTRIALSHUB, we conducted an initial user study with

Model	Tokens	Time	Length
Gemini-3-Pro	1318.44	14.33	698.01
GPT-5.1	227.71	2.81	1006.82
Claude-4.5-Sonnet	210.95	5.34	896.54

Table 4: Performance comparison. **Tokens**: Avg. completion tokens; **Time**: Avg. latency (s); **Length**: Avg. output characters.

seven medical professionals: hematologists, pathologists, dentists, clinical statisticians, pharmacists, pharmaceutical scientists, and nursing researchers. Participants explored the system for tasks corresponding to their typical use of PM and CTG. Table 5 summarizes key findings. Detailed study materials are provided in Appendix D.

Metric	Score
<i>Search Feature Experience (0–5):</i>	
Query generation rating	4.50
Filtering rating	4.40
Eligibility check rating	4.50
Combined search utility	4.14
<i>Information Extraction Accuracy (0–5):</i>	
Study Overview	5.0
Study Plan	5.0
Participation Requirements	4.83
Baseline Characteristics	4.83
Outcome Measures	4.83
Participant Flow	4.7
Adverse Events	4.6
Chatbot answer quality (0–5)	4.86

Table 5: User study results summary. Scores reflect participant ratings of ClinicalTrialsHub functionality.

**Search Stage.** Six of seven participants found 6 or more relevant studies with CLINICALTRIALSHUB compared to 5/7 for PubMed and 3/7 for CTG among the top 30 results, demonstrating improved relevance through unified search with BM25 reranking. Query generation (4.50), filtering (4.40), and eligibility checking (4.50) features received high ratings. The overall combined search capability received strong approval (4.14), validating the value of eliminating cross-platform navigation. Time-saving received moderate rating (3.71), potentially reflecting initial learning curve.

**Review Stage.** Six of seven participants reviewed the structured extraction from PubMed. Perceived accuracy was consistently high: Study Overview and Study Plan received perfect scores (5.0), while Participation Requirements, Baseline Characteristics, and Outcome Measures scored 4.83. Results-oriented sections (Participant Flow:

4.7, Adverse Events: 4.6) also rated highly. All seven participants used the chatbot, rating answer quality (4.86) and overall detail page efficiency (4.86) very highly.

## 7 Related Work

Prior work has aimed to improve clinical trial information access and evidence synthesis in various ways. Trialstreamer (Marshall et al., 2020) structures PubMed articles for rapid evidence browsing. RobotReviewer (Marshall et al., 2016) automates trial data extraction and risk-of-bias evaluation specifically for systematic reviews. LinkedCT (Hassanzadeh et al., 2009) transforms ClinicalTrials.gov data into structured linked data though it does not integrate literature sources, unlike our work. Prior work studied the use of LLMs this space as well—for patient-trial matching, clinical trial design, and participant recruitment (Wang et al., 2024). Most recently, TrialPanorama (Wang et al., 2025a) established a large-scale database and benchmark for these tasks, while LEADS (Wang et al., 2025b) introduced a foundation model specifically designed to enhance human-AI collaboration in medical literature mining. CLINICALTRIALSHUB extends these approaches into a unified platform, bridging registry and literature silos. Unlike Trialstreamer or LinkedCT, it integrates both data sources, and compared to RobotReviewer’s narrow focus on systematic reviews, it supports broader interactive exploration and structured retrieval for diverse clinical and research tasks.

## 8 Conclusion and Future Work

We presented CLINICALTRIALSHUB, a unified platform that integrates structured trial registry data from ClinicalTrials.gov with structured information extracted from unstructured PubMed publications using LLMs. Our system enhances access to comprehensive clinical trial information by enabling unified search, structured information extraction and attributed question answering, supporting the diverse needs of patients, clinicians, and researchers. In future work, we will improve our extraction models through domain-adapted LLM fine-tuning and integrate patient-centric features such as eligibility matching and personalized trial summaries. While our initial user study with seven medical professionals demonstrated the platform’s utility across diverse clinical roles, we plan to conduct larger-scale usability studies to validate these

findings and guide further development.

## References

- Anthropic. 2025. *Claude 4.5 sonnet system card*. Technical report, Anthropic.
- Gemini Team, Google. 2025. *Gemini 3: A multimodal family of capable models*. Technical report, Google DeepMind.
- Oktie Hassanzadeh, Anastasios Kementsietsidis, Lipyeow Lim, Renée J Miller, and Min Wang. 2009. *Linkedct: A linked data space for clinical trials*. Technical Report CSRG-596, University of Toronto.
- Alon Jacovi, Andrew Wang, Chris Alberti, Connie Tao, Jon Lipovetz, Kate Olszewska, Lukas Haas, Michelle Liu, Nate Keating, Adam Bloniarz, Carl Saroufim, Corey Fry, Dror Marcus, Doron Kuklansky, Gaurav Singh Tomar, James Swirhun, Jinwei Xing, Lily Wang, Madhu Gurumurthy, and 7 others. 2025. *The facts grounding leaderboard: Benchmarking llms' ability to ground responses to long-form input*. Preprint, arXiv:2501.03200.
- Harold W. Kuhn. 1955. *The hungarian method for the assignment problem*. *Naval Research Logistics Quarterly*, 2(1–2):83–97.
- Iain J Marshall, Joël Kuiper, and Byron C Wallace. 2016. *Robotreviewer: evaluation of a system for automatically assessing bias in clinical trials*. *Journal of the American Medical Informatics Association*, 23(1):193–201. Originally published online 2015.
- Iain J Marshall, Benjamin Nye, Joël Kuiper, Anna Noel-Storr, Rachel Marshall, Rory Maclean, Frank Soboczenski, Ani Nenkova, James Thomas, and Byron C Wallace. 2020. *Trialstreamer: A living, automatically updated database of clinical trial reports*. *Journal of the American Medical Informatics Association*, 27(12):1903–1912.
- OpenAI. 2025. *Gpt-5.1 system card*. Technical report, OpenAI.
- Stephen E Robertson and Steve Walker. 1994. Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval. In *SIGIR’94: Proceedings of the Seventeenth Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, organised by Dublin City University*, pages 232–241. Springer.
- Zifeng Wang, Lang Cao, Benjamin Danek, Qiao Jin, Zhiyong Lu, and Jimeng Sun. 2024. *Accelerating clinical evidence synthesis with large language models*. Preprint, arXiv:2406.17755.
- Zifeng Wang, Qiao Jin, Jiacheng Lin, Junyi Gao, Jathurshan Pradeepkumar, Pengcheng Jiang, Benjamin Danek, Zhiyong Lu, and Jimeng Sun. 2025a. *Tri-alpanorama: Database and benchmark for systematic review and design of clinical trials*. Preprint, arXiv:2505.16097.
- Zifeng Wang, Qiao Jin, Weicheng Ma, and 1 others. 2025b. *Leads: a foundation model for medical literature mining*. *Nature Communications*, 16(1):1–16.
- Nancy L Wilczynski, Kathleen A McKibbon, and Robert B Haynes. 2011. Sensitive clinical queries retrieved relevant systematic reviews as well as primary studies: an analytic survey. *Journal of Clinical Epidemiology*, 64(12):1341–1349.

## A Additional Statistics Details

Our data expansion is calculated as:

$$\text{Expansion}(\%) = \frac{\text{PMC trials without CTG}}{\text{Total PMC trials}} \times 100 \quad (1)$$

This yields expansion rates of 74.2% (specificity: 102,608/138,279) to 93.4% (sensitivity: 898,846/962,774), averaging 83.8%.

## B Reranking and Result Integration Details

We implement BM25-based reranking mechanisms (Robertson and Walker, 1994) to individual search results, followed by deduplication to merge related publications and trials.

**Corpus Construction:** For PubMed documents, we construct searchable text by concatenating: (1) article titles, (2) abstracts (structured or unstructured), (3) author-provided keywords, (4) Medical Subject Headings (MeSH) descriptors, and (5) journal names. For CTG documents, we concatenate: (1) trial titles, (2) condition specifications, (3) brief summaries, and (4) keywords.

**BM25 Scoring:** During integration, BM25 scores are computed for each source separately, then min-max normalized to the [0, 1] range and combined with a position-based bonus from the original API ranking. For each document  $d_i$ , the score is computed as:

$$\frac{\text{BM25}(d_i) - \text{BM25}_{\min}}{\text{BM25}_{\max} - \text{BM25}_{\min}} + 0.2 \cdot \frac{N - i}{N}$$

where  $\text{BM25}(d_i)$  represents the raw BM25 score,  $\text{BM25}_{\min}$  and  $\text{BM25}_{\max}$  are the minimum and maximum scores for normalization,  $N$  is the total number of results, and  $i$  is the zero-indexed position in the original API ranking. The constant 0.2 represents the maximum position bonus. When  $\text{BM25}_{\max} = \text{BM25}_{\min}$  (indicating identical scores), the BM25 component is set to 0.0 for all documents, and the final ranking relies entirely on the original API position.

**Bidirectional Deduplication:** We merge PubMed and CTG entries that reference each other.³ Merged pairs receive a relevance bonus.⁴ After rescorining and merging, all entries are sorted by their final scores and shown to the user.

## C Information Extraction Details

### C.1 Modular Prompt System

The extraction framework is organized into 3 main sections corresponding to the CTG API v2.0 structure. We provide the exact prompts in our repository.

- **protocolSection:** Contains basic study information including identification modules (NCT ID, title, sponsors), status modules (enrollment, dates, phases), design modules (study type, allocation, masking), arms and interventions, eligibility criteria, contacts, and locations.
- **resultsSection:** Encompasses study outcomes and results when available, including participant flow, baseline characteristics, outcome measures, adverse events, and statistical analyses.
- **derivedSection:** Includes system-generated data elements such as condition browse modules, intervention browse modules, and MeSH term mappings.

### C.2 Data Validation and Schema Conformance

During extraction, we asynchronously operate a validation pipeline.

- **Schema-level validation** enforces structural integrity across all five data sections. Each extracted field undergoes type checking against CTG's built-in types, including character limits for text fields (briefTitle: 300 chars, officialTitle: 600 chars, eligibilityCriteria: 20,000 chars), ISO 8601 date formatting, and proper nesting for complex structures like baseline characteristics and outcome measures.
- **Enumerated value validation** ensures all categorical fields contain only permitted values. The system validates against ClinicalTrials.gov's comprehensive enum definitions for critical fields

³This merging criterion is designed for high precision. We acknowledge that might miss certain trials that may have unidirectional references but do not reference each other (for instance, trials that were registered before the publication but never updated)

⁴We empirically determine this bonus to 0.3.

including studyType, allocation, interventionModel, phases, sex, and standardized age groups. Non-conforming values trigger immediate correction through fuzzy matching against allowed values.

• **Clinical terminology verification** integrates with the NCBI MeSH API to validate medical subject headings in real-time. The system performs fuzzy matching for condition and intervention terms, automatically suggesting and applying standardized MeSH descriptors. This prevents terminology drift and ensures compatibility with biomedical databases.

• **Statistical consistency checking** validates the coherence of quantitative data, verifying group-measure associations in baseline characteristics, confirming unit consistency across measurements, and validating statistical parameters (means, standard deviations, confidence intervals) for mathematical soundness.

Missing information is handled through deliberate omission rather than placeholder generation, preserving data integrity for downstream analysis.

### C.3 Sample ID Filter Query - PubMed

```
(( "randomized controlled trial"[
  Publication Type]
OR "controlled clinical trial"[
  Publication Type]
OR "randomized"[Title/Abstract]
OR "placebo"[Title/Abstract]
OR "clinical trials as topic"[
  MeSH Terms:noexp]
OR "randomly"[Title/Abstract]
OR "trial"[Title])
NOT ("animals"[MeSH Terms] NOT "
  humans"[MeSH Terms]))
AND ("english"[Language] OR "
  English"[lang])
AND "pubmed pmc open access"[
  Filter]
AND "clinicaltrials gov"[
  Secondary Source ID]
AND "2021/01/01" [Date -
  Publication] : "3000" [Date -
  Publication]
```

### C.4 Semantic Similarity Prompt

Compare these clinical trial field values for '{
--------------------------------------------------

```

    field_explanation}'. Return
    only a number 0-1 for semantic
    similarity.
Text1: {reference_value}
Text2: {predicted_value}

```

## D User Study Details

To assess the practical utility of ClinicalTrialsHub in realistic workflows, we conducted an initial user study with seven medical professionals. Participants were recruited across diverse roles: hematology, pathology, dentistry, clinical statistics, pharmacy, pharmaceutical science (PhD student), and nursing (PhD candidate). The participant pool comprised a balanced distribution of clinicians (3/7), clinical researchers (2/7), and students or trainees (2/7), thereby representing key stakeholder groups who utilize clinical trial data. Each participant received an explanation of the system's features and was asked to use CLINICALTRIALSHUB for tasks that correspond to their typical use of PM and CTG. The contexts in which participants explored baseline systems (PM/CTG) and CLINICALTRIALSHUB, along with their stated purposes, are described in [Table 7](#), while the actual queries they input for each platform are shown in [Table 9](#).

### D.1 Search Stage Evaluation

**Query Generation** Although only two participants explicitly used the natural language query generation feature, both rated it highly (4.50). This low usage rate is unsurprising given that most participants were already proficient in constructing Boolean queries for PubMed and CTG. [Table 8](#) shows their natural language inputs and the system-generated structured queries. These results demonstrate that the system accurately decomposes complex natural language requests into structured parameters, particularly benefiting less experienced users or those working outside their primary domain.

**Filtering** Five participants used the filtering capabilities and rated this feature highly (4.40). The filters applied spanned temporal constraints (publication date ranges from 5-10 years), study design parameters (study type—interventional/observational, RCT status, data availability requirements (PMC Open Access, CTG with results posted), and population specifications (humans, age restrictions, completion status). [Table 10](#) documents the spe-

cific filter combinations each participant employed. The CLINICALTRIALSHUB's filtering mechanism enabled users to impose identical selection criteria on heterogeneous data sources, reducing the need to mentally translate filter semantics between platform-specific interfaces.

**Eligibility Check** Four participants utilized the eligibility criteria specification feature, which also received a high average rating (4.50). This feature allowed users to define detailed inclusion and exclusion criteria that went beyond simple filter combinations. For instance, one participant specified inclusion criteria requiring "randomized controlled or single-arm registration trials with  $\geq 50$  patients" that "reported at least one primary outcome," while explicitly excluding "studies including other myeloproliferative neoplasms without fibrosis." A diabetes researcher defined even more granular criteria spanning *intervention type* ("dyadic or family-based behavioral, psychoeducational, or self-management intervention"), *population characteristics* ("adults  $\geq 18$  years with type 2 diabetes"), *outcome requirements* (diabetes self-efficacy, self-management behaviors, dyadic processes, or HbA1c), and *temporal constraints* ("follow-up of at least 3 months"). The complete set of criteria specifications is provided in [Table 11](#). The high rating reflects users' need for nuanced eligibility assessment that cannot be captured through simple keyword filtering alone—particularly for systematic review preparation and evidence synthesis where precise population and study design specifications are critical.

**Overall System Utility** [Table 6](#) shows the distribution of relevant studies participants identified among the top 30 results. 6 of 7 participants found 6 or more relevant studies with CLINICALTRIALSHUB, compared to 5 of 7 for PubMed and only 3 of 7 for CTG. Although CTH demonstrated improved relevance, the maximum response option of '11+' prevents precise quantification of the magnitude of improvement. Nevertheless, our unified search with BM25 reranking successfully integrated heterogeneous data sources into a single ranking system without degrading user satisfaction. This suggests that reducing manual screening effort through cross-source integration is achievable even when reconciling disparate ranking algorithms. The overall combined search capability received strong approval (4.14), validating the value proposition of eliminating cross-platform

Platform	0	1–5	6–10	11+
PubMed	0	2	2	3
ClinicalTrials.gov	0	4	2	1
ClinicalTrialsHub	0	1	3	3

Table 6: Distribution of relevant studies identified among top 30 results.

navigation and manual deduplication. However, the time-saving metric received a more moderate rating (3.71). This may reflect the initial learning curve associated with a new interface or, alternatively, that experienced users already possess efficient workflows for rapidly applying trial filters on baseline systems. We view these responses as an opportunity to identify superior interaction patterns from existing systems and either integrate them into ClinicalTrialsHub or better expose our system’s capabilities to accelerate clinical trial research activities.

## D.2 Review Stage Evaluation

In the review stage, we focused on how well the detail page supported close reading of individual trials. Participants were first asked to rate the accuracy of seven structured sections distilled from PMC (Study Overview, Participation Requirements, Study Plan, Participant Flow, Baseline Characteristics, Outcome Measures, Adverse Events) on a 0–5 scale, then to judge whether this representation and the integrated chatbot detail page helped them interpret studies more efficiently and save time.

**Information Extraction** 6 of 7 participants reported reviewing the structured extraction from PubMed within CLINICALTRIALSHUB. Across those respondents, perceived accuracy was consistently high: both the *Study Overview* and *Study Plan* modules received perfect mean scores of 5.0/5.0, while *Participation Requirements*, *Baseline Characteristics*, and *Outcome Measures* clustered tightly of 4.83. Even the more detailed results-oriented sections, *Participant Flow* (4.7) and *Adverse Events* (4.6, with one non-response), were rated near the top of the scale. These scores suggest that, for the protocol-level fields we benchmark in §4, clinicians also subjectively experience the structured representations as faithful to the source.

**QA** All seven participants used the chatbot on the detail page, and they were asked to provide up

to three concrete examples (paper ID, question, and answer) from their own sessions. Their questions illustrate how the assistant is used as an interpretive layer rather than a generic Q&A tool. Several participants asked design and endpoint focused questions, such as summarizing how the sample size was determined, clarifying what the primary endpoint was and whether it was met, or checking whether specific biomarkers were collected. Others queried safety and practical implications, including requests for the most common adverse events, recommended safest order for clinical steps, or whether particular clinical outcomes showed improvement. Some participants targeted fields in protocol section, such as total enrollment numbers, completion dates, or the geographic distribution of trial sites. Across these diverse uses, participants rated the chatbot’s answers as highly accurate and relevant (4.86) and also agreed that the combined detail page—full text, structured view, and chatbot—helped them interpret study information efficiently (4.86).

Table 7: Participant occupations and task contexts

ID	Occupation	Task for This Evaluation
1	MD, Specializing in Hematology	I am the page editor for the Myelofibrosis evidence page on a hematology platform that aggregates and curates high-quality data from recent hematologic trials. My responsibility is to regularly review newly published clinical trials and update the page with evidence that meets my inclusion criteria. For this evaluation, I need to identify and organize the conclusions of recent myelofibrosis trials that should be added to the Myelofibrosis evidence page.
2	Pharmacist	I am looking for dosing decision evidence for a specific patient who is receiving cefepime and continuous renal replacement therapy (CRRT) in the setting of acute kidney failure, to ensure that the cefepime dose remains below the toxic range while still being effective.
3	PhD Student – pharmaceutical science	I am preparing my candidacy proposal and reviewing preliminary studies exploring aspirin for the prevention of pre-eclampsia in high-risk pregnancies. I would like to examine how prior studies have evaluated placental biomarkers in this context.
4	Statistician	I am designing a clinical trial for a new EGFR-targeted therapy in non-small cell lung cancer. For this evaluation, I need to review existing evidence on EGFR-targeted therapies, including their efficacy and safety profiles, to support the study design.
5	Nurse (PhD Candidate)	I am conducting a systematic review of dyadic and family-based interventions for adults with type 2 diabetes, focusing on diabetes self-management, self-efficacy, and related psychosocial outcomes. My responsibility is to identify recent randomized and registration trials, screen them against predefined inclusion criteria, and organize the conclusions with respect to dyadic or family-based diabetes self-management interventions.
6	Dentist; General Dentistry	I am preparing a paper that discusses the need to standardize IV moderate sedation training and competency assessment in dental residency programs. For this evaluation, I want to identify clinical trials and observational studies on IV moderate sedation in dentistry, including evidence on patient selection, resident training, and the appropriate order and quantity of medications for successful and safe IV moderate sedation.
7	Pathologist (Dermatopathology / Gynecologic Pathology)	I recently reviewed a case involving a metastatic cutaneous melanoma with a BRAFV600E mutation, confirmed through our molecular service. The patient has completed wide excision and sentinel lymph node evaluation, and the oncology team is now considering adjuvant systemic therapy options. For this case, I would like to review the clinical evidence supporting commonly used adjuvant therapies for metastatic melanoma so I can provide informed context when discussing the pathology findings with the treating oncologists. In particular, I want to find studies evaluating adjuvant systemic treatments used after complete resection of metastatic melanoma, including outcomes such as recurrence risk and treatment-related toxicity.

Table 8: Participant Natural Language Queries and Generated Structured Queries

ID	Natural Language Query	Structured Query (CTH Output)
2	Randomized controlled trials involving continuous renal replacement therapy and cefepime in patients with acute kidney injury	Condition: acute kidney injury Intervention: continuous renal replacement therapy OR cefepime Other terms: randomized controlled trials
7	I want clinical trials evaluating chemotherapy in high-stage melanoma, particularly those reporting BRAFV600E mutation status.	Condition: high-stage melanoma OR BRAFV600E mutation Intervention: chemotherapy

Table 9: Participant Search Queries Across Systems

ID	PubMed Query	ClinicalTrials.gov Query	CTH Query
1	myelofibrosis OR polycythemia vera OR essential thrombocythemia OR myeloproliferative	Condition: myelofibrosis OR polycythemia vera OR essential thrombocythemia OR myeloproliferative	Condition: myelofibrosis OR polycythemia vera OR essential thrombocythemia OR myeloproliferative
2	cefepime and crrt	Condition: Acute Kidney Injury (AKI) Intervention: Continuous Renal Replacement Therapy (CRRT) Other terms: cefepime	-
3	Aspirin AND Pre-eclampsia AND biomarkers	Condition: Pre-eclampsia Intervention: aspirin Other terms: biomarkers	Common: pregnancy Condition: pre-eclampsia Intervention: aspirin Other terms: biomarkers, placental biomarkers
4	EGFR-targeted therapies non small cell lung cancer	Condition: non small cell lung cancer Intervention: EGFR-targeted therapies	Common: "EGFR-targeted therapies non small cell lung cancer" Intervention: EGFR-targeted therapies
5	"type 2 diabetes" OR "type 2 diabetes mellitus" OR T2DM AND (dyadic OR family-based OR spouse OR partner OR dyadic OR caregiver OR self-management OR education)	Condition: Type 2 Diabetes Mellitus (T2DM) Other terms: dyadic / family-based interventions, self-management outcomes	Condition: type 2 diabetes mellitus Other terms: dyadic, family-based, spouse, partner, caregiver, self-management, education
6	IV moderate sedation AND midazolam AND fentanyl	Other terms: IV moderate sedation AND midazolam AND fentanyl	Other terms: Dental Anxiety AND Dental procedures AND IV moderate sedation guidelines
7	metastatic melanoma and BRAFV600E mutation and chemotherapy	Condition: metastatic melanoma Intervention: chemotherapy Other terms: BRAF V600E mutation positive	Common: cutaneous melanoma OR adjuvant melanoma OR resected melanoma OR "stage III melanoma" OR "BRAF V600E melanoma"

Table 10: Participant Filters Used Across Systems

ID	PubMed Filters	CTG Filters	CTH Filters
1	Publication Date (Last 5 years) Phase (III/IV) Article Type (Clinical Trial, RCT) Access (PMC OA) Species (Humans)	Completion Date (Last 5 years) Phase (III/IV) Study Type (Interventional) Has Results (True) Status (Completed)	Date (Last 5 years) Phase (III/IV) Study Type (Clinical Trial, RCT) PubMed (PMC OA, Humans) CTG (With Results, Completed)
2	Age Publication Date	Age Completion Date	Age Date
3	Publication Date	Has Results (True)	Phase Study Type
4	Publication Date (Last 10 years) Article Type (Clinical Trial)	Completion Date (From:11/17/2015)	Date (Last 10 years) Study Type (Clinical Trial)
5	Publication Date (Last 10 years) Article Type (Clinical Trial, RCT) Species (Humans) Age (Adults) Language (English)	Completion Date (Last 10 years) Study Type (Interventional, Randomized) Status (Completed) Has Results (True) Age (Adults)	Date (Last 10 years) Study Type (Clinical Trial, RCT) Age (Adults) PubMed (PMC OA, Humans) CTG (With Results, Completed)

Table 11: Participant Inclusion and Exclusion Criteria

<b>ID</b>	<b>Inclusion Criteria</b>	<b>Exclusion Criteria</b>
1	Randomized controlled or single-arm registrational trials with $\geq 50$ patients  Reported at least one primary outcome	Other myeloproliferative neoplasms without fibrosis (e.g., PV or ET without MF)  Prefibrotic MF or early myeloproliferative disease without confirmed fibrosis
2	AKI	
	CRRT	
5	Randomized controlled trials or single-arm registrational trials with approximately $\geq 50$ participants  Adults ( $\geq 18$ years) with type 2 diabetes mellitus  Dyadic or family-based behavioral, psychoeducational, or self-management intervention involving partner/spouse/family  Reported at least one primary or key outcome (self-efficacy, self-management behaviors, dyadic/family processes, HbA1c, etc.)  Follow-up $\geq 3$ months after baseline	Studies focusing only on type 1 diabetes, gestational diabetes, or prediabetes  Purely pharmacologic comparisons without behavioral/educational/family components  Pediatric/adolescent-only populations  Case reports, very small pilot studies, qualitative-only studies without quantitative outcomes
6	Dental	General Anesthesia

Table 12: Participant Chatbot Interactions

ID	Chatbot Interaction
1	<p><b>PMC12092244   What was the exact pelabresib dose and schedule?</b>            Pelabresib was administered orally at a starting dose of 125 mg once daily for 14 consecutive days followed by a 7-day break, forming a 21-day cycle. Dose modifications between 50–175 mg once daily were permitted based on adverse events and treatment response.</p> <p><b>PMC12092244   What was the primary endpoint, and was it met?</b>            The primary endpoint was achieving a $\geq 35\%$ reduction in spleen volume from baseline at week 24. This endpoint was met, with a significantly greater proportion of patients in the pelabresib–ruxolitinib arm achieving this reduction compared to the placebo–ruxolitinib arm.</p> <p><b>PMC12092244   What were the most common adverse events?</b>            The most common treatment-emergent adverse events were hematological—primarily thrombocytopenia and anemia. Among non-hematological events, diarrhea was most frequent in the pelabresib–ruxolitinib arm, whereas constipation was most frequent in the placebo–ruxolitinib arm.</p>
2	<p><b>PMC10576861   Since ACORN showed higher neurotoxicity with Cefepime, should we reconsider de-escalating from Piperacillin-Tazobactam to Cefepime solely based on renal concerns?</b>            Based on the ACORN trial findings, reconsidering the practice of de-escalating from Piperacillin-Tazobactam to Cefepime solely for renal protection is warranted. The study demonstrated that there was no significant difference in the highest stage of acute kidney injury or death between the two treatment groups. Furthermore, the trial concluded that Piperacillin-Tazobactam did not increase the incidence of acute kidney injury, whereas Cefepime was associated with a higher incidence of neurological dysfunction.</p> <p><b>PMC9463324   The study showed creatinine elevation but no difference in dialysis or mortality. Does this mean the current KDIGO criteria are too sensitive for antibiotic-induced AKI?</b>            The study suggests that the creatinine-based KDIGO criteria may identify "pseudotoxicity" rather than true renal injury in the context of Vancomycin and Piperacillin-Tazobactam therapy. The authors propose that the observed creatinine elevation is likely due to the inhibition of tubular secretion rather than damage to the kidney parenchyma, a conclusion supported by the lack of corresponding changes in alternative biomarkers (cystatin C and BUN) and the absence of increased risk for dialysis or mortality.</p>
3	<p><b>NCT03893630   Did the study report any differences in placental biomarker levels between the 81mg and 162mg aspirin groups?</b>            Yes, the study reported specific mean changes in placental biomarker levels for the 81mg and 162mg aspirin groups, although statistical significance of the differences was not explicitly detailed in the provided text. For Placental Growth Factor (PIGF), the mean change from baseline to Event 2 was 568.3 pg/ml for the 81mg group and 509.7 pg/ml for the 162mg group. For soluble Intercellular Adhesion Molecule (s-ICAM), the mean change from baseline to Event 2 was 1.6 ng/ml for the 81mg group and 7.9 ng/ml for the 162mg group. Data for IL-6 and TNF were not reported due to insufficient serum samples.</p>
4	<p><b>PMC12139608   Can you summarize how the sample size is determined in this study?</b>            The study powered for $\geq 85\%$ to detect a 25% difference in ORR using one-sided $\alpha=0.025$; planned enrollment was 126 patients.</p> <p><b>PMC10318754   Can you summarize the statistical analysis methods used?</b>            Analyses included chi-square/Fisher tests and QoL correlations with PFS/OS and sponsorship.</p>
	<p><b>PMC9509093   Summarize tabulated results for all endpoints.</b>            Aumolertinib improved PFS (19.3 vs 9.9 mo); ORR 73.8% vs 72.1%; DCR 93.0% vs 96.7%; DoR 18.1 vs 8.3 mo; fewer rash/diarrhea events.</p>
5	<p><b>PMC8266295   Were any clinical outcomes such as HbA1c reported, and if so, what were the main results?</b>            The paper does not report clinical outcomes such as HbA1c changes as part of the results. The study focused on family supporter involvement, support techniques, and caregiving experience rather than direct clinical outcomes like HbA1c.</p>
6	<p><b>PMC5858011   What do they recommend as the safest order?</b>            Midazolam-first caused fewer desaturation events (90–92%).</p>
7	<p><b>NCT01682083   If OS wasn't clearly significant early on, is the decision to use dabrafenib/trametinib driven mostly by the strong RFS benefit?</b>            Yes, the provided results indicate that the primary outcome, Relapse-free Survival (RFS), demonstrated a statistically significant benefit with a hazard ratio of 0.47 and a p-value of $&lt; 0.0001$. In contrast, the Overall Survival (OS) analysis resulted in a p-value of 0.006, which did not meet the pre-specified significance threshold of 0.000019 for the first interim analysis.</p>

## E Extraction Schema Fields

### E.1 Extraction Schema Evaluation Fields

Piece	# Ref	# EV	Gemini	GPT	Claude
DesignObservationalModel	1	7.67	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>
<b>DetailedDescription</b>	<b>64</b>	<b>31.67</b>	<b>1.00</b>	<b>1.00</b>	0.86
StudyType	100	0.00	<b>1.00</b>	<b>1.00</b>	0.85
DesignAllocation	99	0.67	0.99	<b>1.00</b>	0.85
BriefSummary	100	0.00	<b>1.00</b>	<b>1.00</b>	0.83
Condition	100	0.00	<b>1.00</b>	<b>1.00</b>	0.83
DesignInterventionModel	99	0.33	0.98	<b>1.00</b>	0.85
Phase	99	0.67	0.98	<b>1.00</b>	0.85
EnrollmentCount	100	0.00	0.99	0.99	0.84
DesignMasking	99	0.67	0.98	0.99	0.85
NCTId	100	0.00	0.99	<b>1.00</b>	0.82
<b>Keyword</b>	<b>52</b>	<b>44.00</b>	<b>1.00</b>	<b>1.00</b>	0.80
MinimumAge	99	0.67	0.99	0.99	0.77
EligibilityCriteria	100	0.00	0.99	0.99	0.77
HealthyVolunteers	99	1.00	0.99	0.99	0.76
Sex	100	0.00	0.99	0.99	0.77
<b>MaximumAge</b>	<b>54</b>	<b>39.00</b>	0.99	<b>1.00</b>	0.74
StdAge	100	0.00	0.98	0.95	0.76
DesignWhoMasked	59	3.67	0.88	<b>0.94</b>	0.77
BriefTitle	100	0.00	0.71	<b>1.00</b>	0.84
OfficialTitle	100	0.00	0.71	0.81	0.82

Table 13: Average Extra-Valid (EV) case analysis and information extraction performance for each field. # Ref: number of evaluation cases (out of 100) where the field appears in CTG ground-truth. # EV: average cases across models where the field was Extra-Valid (present in PMC but missing from CTG). Values  $\geq 10$  are bolded. Highest F1 per row is bolded. Rows ordered by average F1.

### E.2 Extraction Schema All Fields

Table 14: Comprehensive list of all fields included in our information-extraction schema.

Index	Piece	Field Index
<b>protocolSection.identificationModule</b>		
1	NCTId	nctId
2	OrgStudyId	orgStudyIdInfo.id
3	OrgStudyIdType	orgStudyIdInfo.type
4	OrgStudyIdLink	orgStudyIdInfo.link
5	SecondaryId	secondaryIdInfos.id
6	SecondaryIdType	secondaryIdInfos.type
7	SecondaryIdLink	secondaryIdInfos.link
8	BriefTitle	briefTitle
9	OfficialTitle	officialTitle
10	Acronym	acronym
11	OrgFullName	organization.fullName

Continued on next page

Table 14 – *Continued from previous page*

Index	Piece	Field Index
12	OrgClass	organization.class
<b>protocolSection.descriptionModule</b>		
13	BriefSummary	briefSummary
14	DetailedDescription	detailedDescription
<b>protocolSection.conditionsModule</b>		
15	Condition	conditions
16	Keyword	keywords
<b>protocolSection.designModule</b>		
17	StudyType	studyType
18	PatientRegistry	patientRegistry
19	TargetDuration	targetDuration
20	Phase	phases
21	DesignAllocation	designInfo.allocation
22	DesignInterventionModel	designInfo.interventionModel
23	DesignInterventionModelDescription	designInfo.interventionModelDescription
24	DesignPrimaryPurpose	designInfo.primaryPurpose
25	DesignObservationalModel	designInfo.observationalModel
26	DesignTimePerspective	designInfo.timePerspective
27	DesignMasking	designInfo.maskingInfo.masking
28	DesignMaskingDescription	designInfo.maskingInfo.maskingDescription
29	DesignWhoMasked	designInfo.maskingInfo.whoMasked
30	EnrollmentCount	enrollmentInfo.count
31	EnrollmentType	enrollmentInfo.type
<b>protocolSection.armsInterventionsModule</b>		
32	ArmGroupLabel	armGroups.label
33	ArmGroupType	armGroups.type
34	ArmGroupDescription	armGroups.description
35	ArmGroupInterventionName	armGroups.interventionNames
36	InterventionType	interventions.type
37	InterventionName	interventions.name
38	InterventionDescription	interventions.description
39	InterventionArmGroupLabel	interventions.armGroupLabels
<b>protocolSection.outcomesModule</b>		
40	PrimaryOutcomeMeasure	primaryOutcomes.measure
41	PrimaryOutcomeDescription	primaryOutcomes.description
42	PrimaryOutcomeTimeFrame	primaryOutcomes.timeFrame
43	SecondaryOutcomeMeasure	secondaryOutcomes.measure
44	SecondaryOutcomeDescription	secondaryOutcomes.description
45	SecondaryOutcomeTimeFrame	secondaryOutcomes.timeFrame
46	OtherOutcomeMeasure	otherOutcomes.measure
47	OtherOutcomeDescription	otherOutcomes.description
48	OtherOutcomeTimeFrame	otherOutcomes.timeFrame
<b>protocolSection.eligibilityModule</b>		

Continued on next page

Table 14 – *Continued from previous page*

Index	Piece	Field Index
49	EligibilityCriteria	eligibilityCriteria
50	HealthyVolunteers	healthyVolunteers
51	Sex	sex
52	MinimumAge	minimumAge
53	MaximumAge	maximumAge
54	StdAge	stdAges
55	StudyPopulation	studyPopulation
56	SamplingMethod	samplingMethod
<b>resultsSection.participantFlowModule</b>		
57	FlowPreAssignmentDetails	preAssignmentDetails
58	FlowRecruitmentDetails	recruitmentDetails
59	FlowTypeUnitsAnalyzed	typeUnitsAnalyzed
60	FlowGroupId	groups.id
61	FlowGroupTitle	groups.title
62	FlowGroupDescription	groups.description
63	FlowPeriodTitle	periods.title
64	FlowMilestoneType	periods.milestones.type
65	FlowMilestoneComment	periods.milestones.comment
66	FlowAchievementGroupId	periods.milestones.achievements.groupId
67	FlowAchievementComment	periods.milestones.achievements.comment
68	FlowAchievementNumSubjects	periods.milestones.achievements.numSubjects
69	FlowAchievementNumUnits	periods.milestones.achievements.numUnits
70	FlowDropWithdrawType	periods.dropWithdraws.type
71	FlowDropWithdrawComment	periods.dropWithdraws.comment
72	FlowReasonGroupId	periods.dropWithdraws.reasons.groupId
73	FlowReasonComment	periods.dropWithdraws.reasons.comment
74	FlowReasonNumSubjects	periods.dropWithdraws.reasons.numSubjects
<b>resultsSection.baselineCharacteristicsModule</b>		
75	BaselinePopulationDescription	populationDescription
76	BaselineTypeUnitsAnalyzed	typeUnitsAnalyzed
77	BaselineGroupId	groups.id
78	BaselineGroupTitle	groups.title
79	BaselineGroupDescription	groups.description
80	BaselineDenomUnits	denoms.units
81	BaselineDenomCountGroupId	denoms.counts.groupId
82	BaselineDenomCountValue	denoms.counts.value
83	BaselineMeasureTitle	measures.title
84	BaselineMeasureDescription	measures.description
85	BaselineMeasurePopulationDescription	measures.populationDescription
86	BaselineMeasureParamType	measures.paramType
87	BaselineMeasureDispersionType	measures.dispersionType
88	BaselineMeasureUnitOfMeasure	measures.unitOfMeasure
89	BaselineMeasureDenomUnits	measures.denoms.units
90	BaselineMeasureDenomCountGroupId	measures.denoms.counts.groupId
91	BaselineMeasureDenomCountValue	measures.denoms.counts.value

Continued on next page

Table 14 – *Continued from previous page*

Index	Piece	Field Index
92	BaselineClassTitle	measures.classes.title
93	BaselineClassDenomUnits	measures.classes.denoms.units
94	BaselineClassDenomCountGroupId	measures.classes.denoms.counts.groupId
95	BaselineClassDenomCountValue	measures.classes.denoms.counts.value
96	BaselineCategoryTitle	measures.classes.categories.title
97	BaselineMeasurementGroupId	measures.classes.categories.measurements.groupId
98	BaselineMeasurementValue	measures.classes.categories.measurements.value
99	BaselineMeasurementSpread	measures.classes.categories.measurements.spread
100	BaselineMeasurementLowerLimit	measures.classes.categories.measurements.lowerLimit
101	BaselineMeasurementUpperLimit	measures.classes.categories.measurements.upperLimit
<b>resultsSection.outcomeMeasuresModule</b>		
102	OutcomeMeasureType	outcomeMeasures.type
103	OutcomeMeasureTitle	outcomeMeasures.title
104	OutcomeMeasureDescription	outcomeMeasures.description
105	OutcomeMeasurePopulationDescription	outcomeMeasures.populationDescription
106	OutcomeMeasureReportingStatus	outcomeMeasures.reportingStatus
107	OutcomeMeasureAnticipatedPostingDate	outcomeMeasures.anticipatedPostingDate
108	OutcomeMeasureParamType	outcomeMeasures.paramType
109	OutcomeMeasureDispersionType	outcomeMeasures.dispersionType
110	OutcomeMeasureUnitOfMeasure	outcomeMeasures.unitOfMeasure
111	OutcomeMeasureCalculatePct	outcomeMeasures.calculatePct
112	OutcomeMeasureTimeFrame	outcomeMeasures.timeFrame
113	OutcomeMeasureTypeUnitsAnalyzed	outcomeMeasures.typeUnitsAnalyzed
114	OutcomeMeasureDenomUnitsSelected	outcomeMeasures.denomUnitsSelected
115	OutcomeGroupId	outcomeMeasures.groups.id
116	OutcomeGroupTitle	outcomeMeasures.groups.title
117	OutcomeGroupDescription	outcomeMeasures.groups.description
118	OutcomeDenomUnits	outcomeMeasures.denoms.units
119	OutcomeDenomCountGroupId	outcomeMeasures.denoms.counts.groupId
120	OutcomeDenomCountValue	outcomeMeasures.denoms.counts.value
121	OutcomeClassTitle	outcomeMeasures.classes.title
122	OutcomeClassDenomUnits	outcomeMeasures.classes.denoms.units
123	OutcomeClassDenomCountGroupId	outcomeMeasures.classes.denoms.counts.groupId
124	OutcomeClassDenomCountValue	outcomeMeasures.classes.denoms.counts.value
125	OutcomeCategoryTitle	outcomeMeasures.classes.categories.title
126	OutcomeMeasurementGroupId	outcomeMeasures.classes.categories.measurements.groupId
127	OutcomeMeasurementValue	outcomeMeasures.classes.categories.measurements.value
128	OutcomeMeasurementSpread	outcomeMeasures.classes.categories.measurements.spread
129	OutcomeMeasurementLowerLimit	outcomeMeasures.classes.categories.measurements.lowerLimit
130	OutcomeMeasurementUpperLimit	outcomeMeasures.classes.categories.measurements.upperLimit
131	OutcomeMeasurementComment	outcomeMeasures.classes.categories.measurements.comment
132	OutcomeAnalysisParamType	outcomeMeasures.analyses.paramType
133	OutcomeAnalysisParamValue	outcomeMeasures.analyses.paramValue
134	OutcomeAnalysisDispersionType	outcomeMeasures.analyses.dispersionType
135	OutcomeAnalysisDispersionValue	outcomeMeasures.analyses.dispersionValue
136	OutcomeAnalysisStatisticalMethod	outcomeMeasures.analyses.statisticalMethod

Continued on next page

Table 14 – *Continued from previous page*

Index	Piece	Field Index
137	OutcomeAnalysisStatisticalComment	outcomeMeasures.analyses.statisticalComment
138	OutcomeAnalysisPValue	outcomeMeasures.analyses.pValue
139	OutcomeAnalysisPValueComment	outcomeMeasures.analyses.pValueComment
140	OutcomeAnalysisCINumSides	outcomeMeasures.analyses.ciNumSides
141	OutcomeAnalysisCIPctValue	outcomeMeasures.analyses.ciPctValue
142	OutcomeAnalysisCILowerLimit	outcomeMeasures.analyses.ciLowerLimit
143	OutcomeAnalysisCIUpperLimit	outcomeMeasures.analyses.ciUpperLimit
144	OutcomeAnalysisCILowerLimitComment	outcomeMeasures.analyses.ciLowerLimitComment
145	OutcomeAnalysisCIUpperLimitComment	outcomeMeasures.analyses.ciUpperLimitComment
146	OutcomeAnalysisEstimateComment	outcomeMeasures.analyses.estimateComment
147	OutcomeAnalysisTestedNonInferiority	outcomeMeasures.analyses.testedNonInferiority
148	OutcomeAnalysisNonInferiorityType	outcomeMeasures.analyses.nonInferiorityType
149	OutcomeAnalysisNonInferiorityComment	outcomeMeasures.analyses.nonInferiorityComment
150	OutcomeAnalysisOtherAnalysisDescription	outcomeMeasures.analyses.otherAnalysisDescription
151	OutcomeAnalysisGroupDescription	outcomeMeasures.analyses.groupDescription
152	OutcomeAnalysisGroupId	outcomeMeasures.analyses.groupIds
<b>resultsSection.adverseEventsModule</b>		
153	EventsFrequencyThreshold	frequencyThreshold
154	EventsTimeFrame	timeFrame
155	EventsDescription	description
156	EventsAllCauseMortalityComment	allCauseMortalityComment
157	EventGroupId	eventGroups.id
158	EventGroupTitle	eventGroups.title
159	EventGroupDescription	eventGroups.description
160	EventGroupDeathsNumAffected	eventGroups.deathsNumAffected
161	EventGroupDeathsNumAtRisk	eventGroups.deathsNumAtRisk
162	EventGroupSeriousNumAffected	eventGroups.seriousNumAffected
163	EventGroupSeriousNumAtRisk	eventGroups.seriousNumAtRisk
164	EventGroupOtherNumAffected	eventGroups.otherNumAffected
165	EventGroupOtherNumAtRisk	eventGroups.otherNumAtRisk
166	SeriousEventTerm	seriousEvents.term
167	SeriousEventOrganSystem	seriousEvents.organSystem
168	SeriousEventSourceVocabulary	seriousEvents.sourceVocabulary
169	SeriousEventAssessmentType	seriousEvents.assessmentType
170	SeriousEventNotes	seriousEvents.notes
171	SeriousEventStatsGroupId	seriousEvents.stats.groupId
172	SeriousEventStatsNumEvents	seriousEvents.stats.numEvents
173	SeriousEventStatsNumAffected	seriousEvents.stats.numAffected
174	SeriousEventStatsNumAtRisk	seriousEvents.stats.numAtRisk
175	OtherEventTerm	otherEvents.term
176	OtherEventOrganSystem	otherEvents.organSystem
177	OtherEventSourceVocabulary	otherEvents.sourceVocabulary
178	OtherEventAssessmentType	otherEvents.assessmentType
179	OtherEventNotes	otherEvents.notes
180	OtherEventStatsGroupId	otherEvents.stats.groupId
181	OtherEventStatsNumEvents	otherEvents.stats.numEvents

Continued on next page

Table 14 – *Continued from previous page*

Index	Piece	Field Index
182	OtherEventStatsNumAffected	otherEvents.stats.numAffected
183	OtherEventStatsNumAtRisk	otherEvents.stats.numAtRisk
<b>resultsSection.moreInfoModule</b>		
184	LimitationsAndCaveatsDescription	limitationsAndCaveats.description
185	AgreementPISponsorEmployee	certainAgreement.piSponsorEmployee
186	AgreementRestrictionType	certainAgreement.restrictionType
187	AgreementRestrictiveAgreement	certainAgreement.restrictiveAgreement
188	AgreementOtherDetails	certainAgreement.otherDetails
189	PointOfContactTitle	pointOfContact.title
190	PointOfContactOrganization	pointOfContact.organization
191	PointOfContactEMail	pointOfContact.email
192	PointOfContactPhone	pointOfContact.phone
193	PointOfContactPhoneExt	pointOfContact.phoneExt
<b>derivedSection.conditionBrowseModule</b>		
194	ConditionMeshId	meshes.id
195	ConditionMeshTerm	meshes.term
196	ConditionAncestorId	ancestors.id
197	ConditionAncestorTerm	ancestors.term
198	ConditionBrowseLeafId	browseLeaves.id
199	ConditionBrowseLeafName	browseLeaves.name
200	ConditionBrowseLeafRelevance	browseLeaves.relevance
201	ConditionBrowseBranchAbbrev	browseBranches.abbrev
202	ConditionBrowseBranchName	browseBranches.name
<b>derivedSection.interventionBrowseModule</b>		
203	InterventionMeshId	meshes.id
204	InterventionMeshTerm	meshes.term
205	InterventionAncestorId	ancestors.id
206	InterventionAncestorTerm	ancestors.term
207	InterventionBrowseLeafId	browseLeaves.id
208	InterventionBrowseLeafName	browseLeaves.name
209	InterventionBrowseLeafRelevance	browseLeaves.relevance
210	InterventionBrowseBranchAbbrev	browseBranches.abbrev
211	InterventionBrowseBranchName	browseBranches.name

## F QA Evaluation Details

Judge Model	Prompt Template	Macro-F1	Acc.	FPR	FNR	F1 (+)	F1 (-)
Claude 3.5 Sonnet	Span-level	68.85	77.83	20.97	22.38	85.58	52.13
	<b>Implicit span-level</b>	<b>70.24</b>	83.50	45.16	11.34	90.10	50.37
	Response-level	61.88	83.25	72.58	6.69	90.42	33.33
	JSON	56.04	64.78	33.87	35.47	75.64	36.44
	JSON (alt)	55.37	66.75	46.77	30.81	77.91	32.84
	JSON w. double-check	49.50	54.68	25.81	48.84	65.67	33.33
Gemini 1.5 Pro	SimpleQA template	55.39	85.22	88.71	1.45	91.87	18.92
	Span-level	55.84	79.31	79.03	10.17	88.03	23.64
	Implicit span-level	56.66	85.47	87.10	1.45	91.99	21.33
	Response-level	48.82	82.02	95.16	4.07	90.04	7.59
	<b>JSON</b>	<b>71.47</b>	86.95	56.45	5.23	92.48	50.47
	JSON (alt)	66.03	85.96	69.35	4.07	92.05	40.00
GPT-4o	JSON w. double-check	64.89	76.35	37.10	21.22	84.95	44.83
	SimpleQA template	51.54	84.73	93.55	1.16	91.64	11.43
	Span-level	63.08	81.53	64.52	10.17	89.18	36.97
	Implicit span-level	55.43	83.99	87.10	3.20	91.11	19.75
	Response-level	51.54	84.73	93.55	1.16	91.64	11.43
	<b>JSON</b>	<b>69.68</b>	80.54	32.26	17.15	87.83	51.53
Qwen 4.0	JSON (alt)	66.78	82.02	53.23	11.63	89.28	44.27
	JSON w. double-check	57.62	64.04	17.74	39.24	74.11	41.13
	SimpleQA template	47.04	83.74	98.39	1.45	91.13	2.94

Table 15: Evaluation of judge models and prompt templates, reproduced from the **FACTS Grounding Benchmark** paper (Table 2).

## G User Interface - Search Page

The screenshot shows the Clinical Trials Hub search interface. At the top, there are tabs for 'Home' and 'About'. Below the header, the title 'Clinical Trials Hub' is centered. A search bar contains the query 'type2 diabetes insulin child'. To the right of the search bar is a magnifying glass icon and a gear icon. The main search area has three sections: 'Condition' (Diabetes), 'Intervention' (Insulin), and 'Other Terms' (RCT). Below these are two boxes: 'PubMed Query' (Myelofibrosis AND Randomized Controlled Trial[PT]) and 'CTG Query' (Hypertension AND ("Cleveland, Ohio") AND recruiting). Underneath the search bar, it says 'Select Browsing Mode: Expert ✓ Patient'. The results section is titled 'Search Results' and contains two parts: 'PubMed Search Query' with the query '((type 2 diabetes) AND (insulin) AND (child)) AND pubmed pmc open access[Filter]' and 'ClinicalTrials.gov Search Query' with the condition 'type 2 diabetes | Intervention: insulin | Other terms: child'. At the bottom, it shows 'Total: 1285 Merged: 2 PubMed-only: 998 CTG-only: 285'.

Figure 4: Search panel

The screenshot shows the Advanced Search Builder panel. At the top, there are tabs for 'Home' and 'About'. The main area is titled 'Advanced Search Builder'. It includes a 'Search Term' input field with placeholder 'Enter term or select from below...', an 'AND' button, an 'OR' button, a 'NOT' button, and parentheses buttons. To the right is a 'Query Box' containing the query '(type 2 diabetes OR diabetes mellitus type 2 OR type II diabetes) AND insulin AND (child OR pediatric OR children OR childhood OR minor)' with a 'Clear' button. Below this is a 'Search' button. The results preview shows 'PubMed: 996 CTG: 964 Merged: 4 Total: 1964' and a link to 'View Full Results'. The panel is divided into sections: 'Refined Condition' (type 2 diabetes + diabetes mellitus type 2 + adult-onset diabetes + non-insulin-dependent diabetes mellitus + type II diabetes), 'Refined Intervention' (insulin + recombinant insulin + human insulin + insulin therapy + insulin analogs), and 'Refined Other Term' (child + pediatric + children + childhood + minor). Each section shows its total count: 1964, 1961, and 1963 respectively.

Figure 5: Advanced search panel

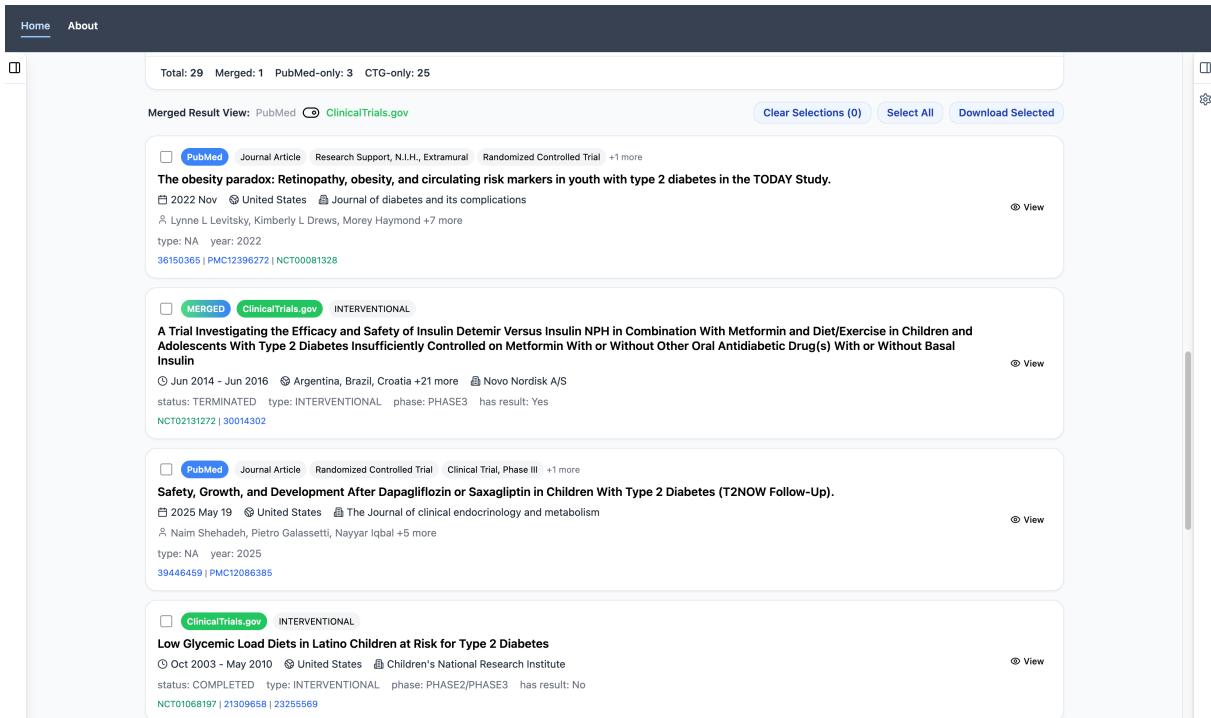


Figure 6: Search results

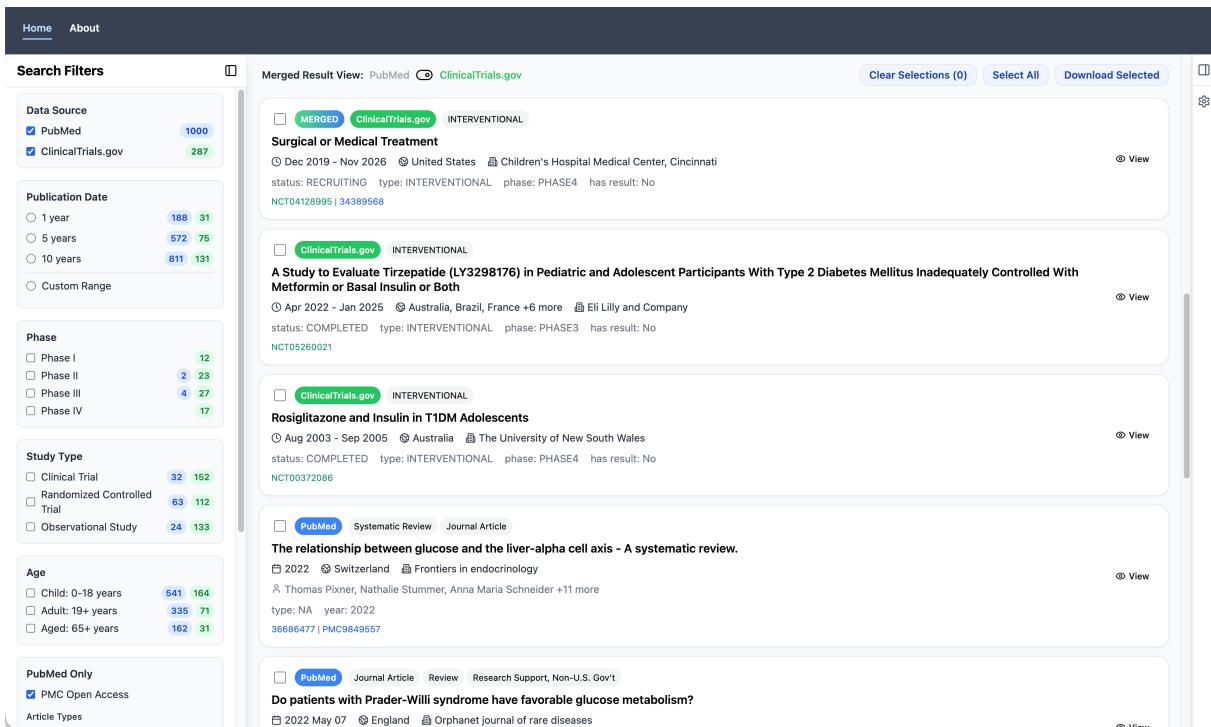


Figure 7: Filtering sidebar

Home About

Search Results

Total: 1285 Merged: 2 PubMed-only: 998 CTG-only: 285

Merged Result View: PubMed ClinicalTrials.gov Clear Selections (0) Select All Download Selected

MERGED ClinicalTrials.gov INTERVENTIONAL  
**Surgical or Medical Treatment**  
 ☰ Dec 2019 - Nov 2026 ☰ United States ☰ Children's Hospital Medical Center, Cincinnati  
 status: RECRUITING type: INTERVENTIONAL phase: PHASE4 has result: No  
 NCT04128995 | 34389568

ClinicalTrials.gov INTERVENTIONAL  
**A Study to Evaluate Tirzepatide (LY3298176) in Pediatric and Adolescent Participants With Type 2 Diabetes Mellitus Inadequately Controlled With Metformin or Basal Insulin or Both**  
 ☰ Apr 2022 - Jan 2025 ☰ Australia, Brazil, France +6 more ☰ Eli Lilly and Company  
 status: COMPLETED type: INTERVENTIONAL phase: PHASE3 has result: No  
 NCT05260021

ClinicalTrials.gov INTERVENTIONAL  
**Rosiglitazone and Insulin in T1DM Adolescents**  
 ☰ Aug 2003 - Sep 2005 ☰ Australia ☰ The University of New South Wales  
 status: COMPLETED type: INTERVENTIONAL phase: PHASE4 has result: No  
 NCT00372086

PubMed Systematic Review Journal Article  
**The relationship between glucose and the liver-alpha cell axis - A systematic review.**  
 ☰ 2022 ☰ Switzerland ☰ Frontiers in endocrinology  
 ↗ Thomas Pixner, Nathalie Stummer, Anna Maria Schneider +11 more  
 type: NA year: 2022  
 36886477 | PMC9849557

ClinicalTrials.gov Preview

NCTID NCT02131272  
 Related PMIDs 30014302

Study Design

Study Type	INTERVENTIONAL
Phase	PHASE3
Enrollment	42 (ACTUAL)
Conditions	Diabetes Diabetes Mellitus, Type 2
Interventions	Diet/exercise Insulin detemir Insulin NPH
Arm Groups	EXPERIMENTAL: Insulin detemir and diet/exercise ACTIVE COMPARATOR: Insulin NPH and diet/exercise

Eligibility Criteria

Age Groups	CHILD, ADULT
Sexes	ALL
Age Range	10 Years to 18 Years
Healthy Volunteers	No

Primary Outcomes

- Change in HbA1c (Glycosylated Haemoglobin)

Secondary Outcomes

- Change in Body Weight Standard Deviation Score (SDS)
- Proportion of Subjects Achieving HbA1c Below 7.0%, Who Have Not Experienced Any Treatment Emergent Severe Hypoglycaemic Episodes Within the Last 14 Weeks of Treatment.
- Proportion of Subjects Achieving HbA1c Below 7.5%, Who Have Not Experienced Any Treatment Emergent Severe Hypoglycaemic Episodes Within the Last 14 Weeks of Treatment

Figure 8: Preview sidebar

Home About

Search Filters

Data Source  
 PubMed 4  
 ClinicalTrials.gov 26

Publication Date  
 1 year 1 4  
 5 years 2 11  
 10 years 4 19  
 Custom Range

Phase  
 Phase I 0  
 Phase II 4  
 Phase III 4 26  
 Phase IV 0

Study Type  
 Clinical Trial 4 26  
 Randomized Controlled Trial 4 26  
 Observational Study 0

Age  
 Child: 0-18 years 4 20  
 Adult: 19+ years 1 4  
 Aged: 65+ years 1 1

PubMed Only  
 PMC Open Access  
 Article Types

ClinicalTrials.gov Preview

Rationale, design, and methods for the Medical Optimization and Management of Pregnancies with Overt Type 2 Diabetes (MOMPOD) study.  
 ☰ 2018 Dec 12 ☰ England ☰ BMC pregnancy and childbirth  
 ↗ Diane C Berry, Sonia Davis Thomas, Karen F Dorman +4 more  
 type: NA year: 2018  
 30541506 | PMC6292086 | NCT02932475

MERGED ClinicalTrials.gov INTERVENTIONAL  
**A Trial Investigating the Efficacy and Safety of Insulin Detemir Versus Insulin NPH in Combination With Metformin and Diet/Exercise in Children and Adolescents With Type 2 Diabetes Insufficiently Controlled on Metformin With or Without Other Oral Antidiabetic Drug(s) With or Without Basal Insulin**  
 ☰ Jun 2014 - Jun 2016 ☰ Argentina, Brazil, Croatia +21 more ☰ Novo Nordisk A/S  
 status: TERMINATED type: INTERVENTIONAL phase: PHASE3 has result: Yes  
 NCT02131272 | 30014302

PubMed Journal Article Randomized Controlled Trial Clinical Trial, Phase III +1 more  
**Safety, Growth, and Development After Dapagliflozin or Saxagliptin in Children With Type 2 Diabetes (T2NOW Follow-Up).**  
 ☰ 2025 May 19 ☰ United States ☰ The Journal of clinical endocrinology and metabolism  
 ↗ Naim Shehadeh, Pietro Galassetti, Nayyar Iqbal +5 more  
 type: NA year: 2025  
 39446459 | PMC12086385

ClinicalTrials.gov INTERVENTIONAL  
**Study to Evaluate Safety and Efficacy of Dapagliflozin in Patients With Type 2 Diabetes Mellitus Aged 10-24 Years**  
 ☰ Jun 2016 - Apr 2020 ☰ Hungary, Israel, Mexico +4 more ☰ AstraZeneca  
 status: COMPLETED type: INTERVENTIONAL phase: PHASE3 has result: Yes  
 NCT02725593 | 35378069

Eligibility Assessment

Eligibility Assessment

Eligibility Assessment

Inclusion Criteria

Intervention includes ACE inhibitors (e.g., lisinopril, enalapril, ramipril)  
 ↗ Not Met

Truth: False Confidence: 90%

Evidence

Interventions: DRUG: Dapagliflozin, DRUG: Dapagliflozin placebo

Reasoning

The abstract explicitly mentions the interventions as dapagliflozin and its placebo. There is no mention of ACE inhibitors such as lisinopril, enalapril, or ramipril.

Outcomes include HbA1c, fasting glucose, or similar  
 ↗ Met biomarkers

Truth: True Confidence: 90%

Evidence

Primary Outcomes: Adjusted Change From Baseline in Glycated Haemoglobin (HbA1c) at Week 24

Reasoning

The abstract explicitly states that the primary outcome includes HbA1c, which is a biomarker related to diabetes management.

Exclusion Criteria

Non-RCT designs (observational, cohort, etc.)  
 ↗ Not violated

Figure 9: Eligibility check results

The screenshot shows the AI Insights & Analysis section of a clinical trial summary page. At the top, it displays the study title, 'Omnipod Horizon™ Automated Glucose Control System Preschool Cohort', and basic details like the start date (Sep 2020 - Oct 2022), location (United States), sponsor (Insulet Corporation), and status (COMPLETED). A 'View' button is also present.

**AI Insights & Analysis** (10 items analyzed)

**Research Overview**

The current research landscape is focused on diabetes, particularly its management and effects in various populations, including children and pregnant women. Despite the absence of clinical trials, recent studies provide insights into diabetes treatment options and patient experiences, emphasizing the need for further exploration in these areas.

**Key Findings**

- Finding 1: Intramyocellular triglyceride content is being studied in the context of both Type 1 and Type 2 diabetes, highlighting metabolic changes early in the disease course.
- Finding 2: Cognitive and neuropsychological profiles of children exposed to metformin versus insulin for gestational diabetes are being evaluated, indicating a focus on long-term outcomes of prenatal diabetes management.
- Finding 3: Continuous glucose monitoring and sensor-augmented insulin pump therapy are being reviewed for patient experiences, suggesting an interest in qualitative assessments of diabetes technologies.

**Research Trends**

- Trend 1: There is a growing interest in the qualitative aspects of diabetes management, as evidenced by systematic reviews of patient experiences with advanced monitoring technologies.
- Trend 2: Research is increasingly focusing on pediatric populations, particularly in relation to diabetes and prediabetes in children with cystic fibrosis.

**Recommendations**

- Recommendation 1: Researchers should consider conducting clinical trials to validate findings from observational studies, particularly in the areas of diabetes management and interventions.
- Recommendation 2: Further exploration into the long-term cognitive effects of gestational diabetes treatments on children is recommended to guide clinical practices.

**Research Gaps**

- Gap 1: There is a lack of clinical trials assessing the efficacy and safety of newer insulin analogs in young children, which could provide more robust evidence for pediatric diabetes management.
- Gap 2: More research is needed on the psychological impacts of diabetes technologies on patients, as current studies focus primarily on physiological outcomes.

Page 1 of 193

Figure 10: AI Insights

## H User Interface - Detail Page

**ChatBot**

**Q:** Did the trial include participants who were already using basal insulin?

**A:** Yes, the trial included participants who were already using insulin as part of their standard of care. The inclusion criteria specified that eligible participants could be on a stable dose of insulin, either alone or in combination with metformin. Furthermore, the study protocol noted that basal insulin could be up-titrated as a rescue measure, implying its potential presence in the treatment regimen.

**Evidence:**

- Inclusion criteria were HbA1c 6.5–11%, FPG ≤14.2 mmol/L ( $\leq 256$  mg/dL) and a stable dose of either metformin ( $\geq 1000$  mg daily), insulin, or a combination of metformin ( $\geq 1000$  mg daily) plus insulin for a minimum of 8 weeks.
- Basal insulin was initiated or up-titrated as open-label rescue for participants meeting prespecified criteria for lack of glycemic control.

Ask about PMC10851108

**Structured Information**

Protocol Section | Results Section | Document Section | Derived Section

**Participant Flow**

**Recruitment Details**

Participants took part in the study at 42 study centres in 7 countries worldwide.

**Arm/Group Information**

ID	Title	Description
FG000	Dapagliflozin 10mg/ Dapagliflozin 10mg	Dapagliflozin (10 mg) tablet administered orally, once daily for the 24 week double-blinded treatment period. The participants then continued to receive Dapagliflozin (10 mg) once daily for a further 28 weeks in the open label long term-extension.
FG001	Placebo/ Dapagliflozin 10mg	Matching placebo tablet administered orally, once daily for the 24 weeks double-blinded treatment period. The participants then received Dapagliflozin (10 mg), orally, once daily for a further 28 weeks in the open label long-term extension.

**Study Periods**

Discrete stages of a clinical study during which numbers of participants at specific significant events or points of time are reported.

Arm/Group Title	Dapagliflozin 10mg/ Dapagliflozin 10mg	Placebo/ Dapagliflozin 10mg
<b>Blinded Treatment Period</b>		
Started	39	33
Received Treatment	39	33
Completed	34	27
Not Completed	5	6
<b>Long Term Extension</b>		
Started	33	27
Completed	32	24
Not Completed	1	3

**Baseline Characteristics**

A description of each baseline or demographic characteristic measured in the clinical study

**Baseline Groups**

ID	Title	Description
.....	Dapagliflozin 10mg/	Dapagliflozin (10 mg) tablet administered orally, once daily for the 24 week double-blinded treatment

Figure 11: Detail page

**References**

**From PubMed**  
These publications come from PubMed, a public database of scientific and medical articles.

1. Tamborlane WV, Laffel LM, Shehadeh N, Isganaitis E, Van Name M, Ratnayake J, Karlsson C, Norjavaara E. Efficacy and safety of dapagliflozin in children and young adults with type 2 diabetes: a prospective, multicentre, randomised, parallel group, phase 3 study. *Lancet Diabetes Endocrinol.* 2022 May;10(5):341-350. doi: 10.1016/S2213-8587(22)00052-3. Epub 2022 Apr 1.  
PMID: 35378069 | PMCID: PMC10851108  
[Collapse Full Text](#)

**Full Text for Reference 1**

**Randomization and masking**  
Participants were stratified by sex, age (10–15, >15–<18, ≥18–<25 years) and background medication (metformin alone, insulin alone, or insulin+metformin). *A priori* recruitment of participants aged 18–<25 years was limited to <40% of the total population, while recruitment of participants aged 10–15 years was to comprise ≥20% of the total population. An interactive web/voice response system randomly assigned treatment (placebo or study drug) to each participant. During the 24-week efficacy period, participants and study personnel were blinded to treatment. Treatment during the subsequent 28-week extension period was open-label, although blinding with respect to treatment received in the initial 24-week period was maintained. The sponsor was responsible for randomization and blinding.

**Procedures**  
The study treatments were oral, once-daily, dapagliflozin 10 mg or placebo added to standard of care (metformin alone, insulin alone or metformin+insulin). Participants were assessed weekly during the 4-week lead-in period, at baseline, during the double-blind period (Week 1, 2, 4, 8, 10, 12, 16, 20, 24), during the open-label safety extension (Week 28, 32, 36, 40, 46, 52) and 4 weeks post treatment. Basal insulin was initiated or up-titrated as open-label rescue for participants meeting prespecified criteria for lack of glycemic control: FPG >13.3 mmol/L (>240 mg/dL) during the double-blind period; FPG >10 mmol/L (>180 mg/dL) or HbA1c >8.0% during the open-label period ([Table S2](#)). Rescued participants continued treatment with the study drug and continued in the study. Those receiving insulin (as background or as glycemic rescue) underwent dose adjustments as per investigators' discretion.

**Outcome measures**

**Efficacy**  
The primary efficacy outcome was mean change from baseline to 24 weeks in HbA1c with dapagliflozin 10 mg versus placebo. Values after glycemic rescue or permanent discontinuation from study drug were excluded. A prespecified sensitivity analysis was to be performed if >10% of participants in either treatment group had protocol deviations predefined as affecting the primary efficacy results. These predefined protocol deviations are described in [Table S3](#). The primary efficacy endpoint was also described according to subgroups: sex, race (white, non-white), baseline HbA1c (<8%, ≥8%) and background medication (insulin+metformin, metformin only).

Secondary endpoints, in order of hierarchical testing, were mean change from baseline to 24 weeks in FPG, percentage of participants who received glycemic rescue or discontinued study due to lack of glycemic control up to 24 weeks, and percentage of participants with baseline HbA1c ≥7% who achieved HbA1c <7% at 24 weeks.

**Safety**  
Safety and tolerability were assessed throughout and included reporting of adverse events (AEs), serious AEs (SAEs), discontinuation due to AEs, hypoglycemia, diabetes ketoacidosis (adjudicated by committee), hepatic laboratory parameters, and vital signs (height, weight, BMI z-score and blood pressure).

Hypoglycemia was defined according to American Diabetes Association (ADA) criteria: severe (required assistance to administer carbohydrate, glucagon or other actions to promote neurological recovery), documented symptoms (typical symptoms and/or plasma glucose [PG] ≤3.9 mmol/L [ $\le 70 \text{ mg/dL}$ ]), asymptomatic (no symptoms, PG ≤3.9 mmol/L [ $\le 70 \text{ mg/dL}$ ]), probable symptomatic (typical symptoms without a glucose

Figure 12: QA evidence highlight

## I Search Query Generation Prompt Templates

You are a clinical expert skilled in transforming a user's natural language query into precise search queries for PubMed to identify the most relevant clinical trial papers.

The user's input is provided as key-value pairs in JSON format as follows:

```
{  
  "user_query": a free text natural language query,  
  "cond": condition/disease terms,  
  "intr": intervention/treatment terms,  
  "other_term": other terms  
}
```

Return your result in JSON format exactly as follows:

```
{  
  "cond": refined condition/disease terms,  
  "intr": refined intervention/treatment terms,  
  "other_term": refined other terms,  
  "combined_query": the final combined search query created by concatenating the refined 'cond', 'intr', and 'other_te  
rm' using the 'AND' operator with parentheses for non-empty fields.  
}
```

Please refine the user's input by following these rules:

- If the user's input is not in English, first translate it into English.
- Retain only the meaningful keywords in each field.
- The "user_query" field is the primary unstructured input where the user might write, for example, "Find clinical trials for depressive disorder using medication," etc. Extract only the meaningful keywords from "user_query" and assign them to "cond", "intr", or "other_term" as appropriate. In other words, merge the information in "user_query" with any provided in "cond", "intr", and "other_term" so that the final refined values capture all the important concepts.
- If any keywords in "user_query" or "other_term" can be classified as condition or intervention terms, move them to "cond" or "intr" accordingly and remove them from "other_term".
- For each field, if there are multiple keywords, combine them using appropriate Boolean operators (AND, OR) and grouping with parentheses.

Example:

- User's input:  
{  
 "user_query": "Find clinical trials for depressive disorder using medication; also interested in ADHD in children, i  
ncluding studies on methylphenidate",  
 "cond": "depressive disorder",  
 "intr": "medication",  
 "other_term": "atomoxetine"  
}

- Your output should be:

```
{  
  "cond": "depressive disorder OR adhd",  
  "intr": "Methylphenidate OR Atomoxetine",  
  "other_term": "child",  
  "combined_query": "(depressive disorder OR adhd) AND (Methylphenidate OR Atomoxetine) AND (child)"  
}
```

Note: The final search query (the value of "combined_query") will be used for the database search, and the content of "user_query" itself will not be used directly.

Figure 13: Search Query Generation - System Prompt

```

The user's input for clinical trial search:

{{inputData}}


Return your result in JSON format as follows:
{
  "cond": refined condition/disease terms,
  "intr": refined intervention/treatment terms,
  "other_term": refined other terms,
  "combined_query": final combined search query created by concatenating the refined 'cond', 'intr', and 'other_term' using the 'AND' operator with parentheses for non-empty fields.
}

```

Figure 14: Search Query Generation - User Prompt

## J Information Extraction Prompt Templates

```

# CONTEXT #
You are tasked with analyzing clinical trial study reports or papers to extract specific information as structured data. Omit any fields if the corresponding data does not exist in the target articles. Do not invent or assume information that is not present in the given content.

# PAPER CONTENT #
{{pmc_text}}


# TARGET #
Extract the identificationModule of the study. It includes identifying information such as the trial's unique identifier or title, illustrating "who is conducting which trial and where it is registered."


# RESPONSE #
The JSON response must adhere to the following structure, where each field is annotated with its expected type.
Ensure that the output is a valid and correctly formatted JSON string.

Format:
```json
{
  "identificationModule": {
    "nctId": "\\ ClinicalTrials.gov Identifier. The format is \"NCT\" followed by an 8-digit number: TEXT (max 11 chars)
    "orgStudyIdInfo": {
      "id": "\\ organization's unique protocol ID: TEXT (max 30 chars)
      "type": "\\ Type of organization-issued ID: ENUM (NIH, FDA, VA, CDC, AHRQ, SAMHSA)
      "link": "\\ URL link related to OrgStudyId and OrgStudyIdType: TEXT
    },
    "secondaryIdInfos": [ \\ ARRAY of OBJECT
      {{
        "id": "\\ Secondary identifier for funding or registry: TEXT (max 30 chars)
        "type": "\\ Type of secondary ID: ENUM (NIH, FDA, VA, CDC, AHRQ, SAMHSA, OTHER_GRANT, EUDRACT_NUMBER, CTIS, REGISTRY, OTHER)
        "domain": "\\ Name of funding organization, registry, or issuer: TEXT (max 119 chars)
        "link": "\\ URL link related to SecondaryId and SecondaryIdType: TEXT
      }}
    ],
    "organization": {{
      "fullName": "\\ Name of the sponsoring organization: TEXT
      "class": "\\ Organization type: ENUM (NIH, FED, OTHER_GOV, INDIV, INDUSTRY, NETWORK, AMBIG, OTHER, UNKNOWN)
    }},
    "briefTitle": "\\ Short title of the study: TEXT (max 300 chars)
    "officialTitle": "\\ Full official title of the study: TEXT (max 600 chars)
    "acronym": "\\ Study acronym: TEXT (max 14 chars)
  }
}
```

```

Figure 15: Protocol Section - Identification Module

```

# CONTEXT #
You are tasked with analyzing clinical trial study reports or papers to extract specific information as structured data. Omit any fields if the corresponding data does not exist in the target articles. Do not invent or assume information that is not present in the given content.

# PAPER CONTENT #
{{pmc_text}}


# TARGET #
Extract the descriptionModule and conditionsModule of the study. DescriptionModule offers a brief introduction or summary of the clinical trial. ConditionsModule specifies the target conditions or topics (keywords), indicating which are being studied.

# RESPONSE #
The JSON response must adhere to the following structure, where each field is annotated with its expected type.
Ensure that the output is a valid and correctly formatted JSON string.

Format:
```json
{
  "descriptionModule": {},
    "briefSummary": \\ Concise summary of the study, including its hypothesis, in layman's terms: TEXT (max 5000 chars)
    "detailedDescription": \\ Extended study description with technical details, excluding full protocol or duplicate information: TEXT (max 32000 chars)
  },
  "conditionsModule": {
    "conditions": \\ List of disease(s) or condition(s) studied, preferably using MeSH or SNOMED CT terms: ARRAY of TEXT
    "keywords": \\ List of descriptive words or phrases related to the study, preferably using MeSH terms: ARRAY of TEXT
  }
}
```
...
```

```

Figure 16: Protocol Section - Description Module, Conditions Module

```

# CONTEXT #
You are tasked with analyzing clinical trial study reports or papers to extract specific information as structured data. Omit any fields if the corresponding data does not exist in the target articles. Do not invent or assume information that is not present in the given content.

# PAPER CONTENT #
{{pmc_text}}


# TARGET #
Extract the designModule of the study. It defines the overall study design (study type, phase, allocation, masking, number of participants, etc.) in detail.

# RESPONSE #
The JSON response must adhere to the following structure, where each field is annotated with its expected type.
Ensure that the output is a valid and correctly formatted JSON string.

Format:
```json
{
  "designModule": {},
    "studyType": \\ Study classification: ENUM (EXPANDED_ACCESS, INTERVENTIONAL, OBSERVATIONAL)
    "patientRegistry": \\ Indicates if the study is a patient registry: BOOLEAN
    "targetDuration": \\ Follow-up duration for observational patient registry studies: TIME
    "phases": \\ Study phase(s), applicable for drug/biologic trials: ARRAY of ENUM (NA, EARLY_PHASE1, PHASE1, PHASE2, PHASE3, PHASE4)
    "designInfo": {
      "allocation": \\ Method of assigning participants: ENUM (RANDOMIZED, NON_RANDOMIZED, NA)
      "interventionModel": \\ Type of intervention design: ENUM (SINGLE_GROUP, PARALLEL, CROSSOVER, FACTORIAL, SEQUENTIAL)
      "interventionModelDescription": \\ Description of the intervention model: TEXT
      "primaryPurpose": \\ The main objective of the intervention(s) being evaluated by the clinical trial: ENUM (TREATMENT, PREVENTION, DIAGNOSTIC, EDUCATION, SUPPORTIVE_CARE, SCREENING, HEALTH_SERVICES_RESEARCH, BASIC_SCIENCE, DEVICE_FEASIBILITY, OTHER)
      "observationalModel": \\ Study model for observational studies: ENUM (COHORT, CASE_CONTROL, CASE_ONLY, CASE_CROSSOVER, ECOLOGIC_OR_COMMUNITY, FAMILY_BASED, DEFINED_POPULATION, NATURAL_HISTORY, OTHER)
      "timePerspective": \\ Time perspective for observational studies: ENUM (RETROSPECTIVE, PROSPECTIVE, CROSS_SECTIONAL, OTHER)
      "maskingInfo": {
        "masking": \\ Level of blinding: ENUM (NONE, SINGLE, DOUBLE, TRIPLE, QUADRUPLE)
        "maskingDescription": \\ Detailed description of masking: TEXT (max 1000 chars)
        "whoMasked": \\ Groups involved in masking: ARRAY of ENUM (PARTICIPANT, CARE_PROVIDER, INVESTIGATOR, OUTCOMES_ASSESSOR)
      }
    },
    "enrollmentInfo": {
      "count": \\ Number of participants enrolled: NUMERIC
      "type": \\ Actual or estimated enrollment: ENUM (ACTUAL, ESTIMATED)
    }
}
```
...
```

```

Figure 17: Protocol Section - Design Module

```

# CONTEXT #
You are tasked with analyzing clinical trial study reports or papers to extract specific information as structured data.
# PAPER CONTENT #
{{pmc_text}}


# TARGET #
Extract the armsInterventionsModule of the study. It details which drugs (or procedures) are administered, to which arms (groups), and how they are applied.
# RESPONSE #
A syntactically correct JSON string:
Format:
```json
{
  "armsInterventionsModule": {
    "armGroups": [ \\ ARRAY of OBJECT
      {
        "label": \\ Name of the arm/group: TEXT
        "type": \\ Type of arm: ENUM (EXPERIMENTAL, ACTIVE_COMPARATOR, PLACEBO_COMPARATOR, SHAM_COMPARATOR, NO_INTERVENTION, OTHER)
        "description": \\ Description of the arm/group: TEXT
        "interventionNames": \\ List of interventions used in this arm/group: ARRAY of TEXT
      }
    ],
    "interventions": [ \\ ARRAY of OBJECT
      {
        "type": \\ Type of intervention: ENUM (BEHAVIORAL, BIOLOGICAL, COMBINATION_PRODUCT, DEVICE, DIAGNOSTIC_TEST, DIETARY_SUPPLEMENT, DRUG, GENETIC, PROCEDURE, RADIATION, OTHER)
        "name": \\ Name of the intervention: TEXT
        "description": \\ Description of the intervention: TEXT
        "armGroupLabels": \\ List of arm/group labels associated with this intervention: ARRAY of TEXT
      }
    ]
  }
},
```

```

Figure 18: Protocol Section - Arms Interventions Module

```

# CONTEXT #
You are tasked with analyzing clinical trial study reports or papers to extract specific information as structured data. Omit any fields if the corresponding data does not exist in the target articles. Do not invent or assume information that is not present in the given content.
# PAPER CONTENT #
{{pmc_text}}


# TARGET #
Extract the outcomesModule of the study. It describes the primary, secondary, and other outcome measures, showing which indicators are used to assess the trial's effectiveness and safety.
# RESPONSE #
The JSON response must adhere to the following structure, where each field is annotated with its expected type.
Ensure that the output is a valid and correctly formatted JSON string.
Format:
```json
{
  "outcomesModule": {
    "primaryOutcomes": [ \\ Required. List of primary outcome measures used to assess the trial's main objectives: ARRAY of OBJECT
      {
        "measure": \\ Name of the primary outcome measure: TEXT (max 255 chars)
        "description": \\ Description of the metric used to characterize the primary outcome measure: TEXT (max 999 chars)
        "timeFrame": \\ Time point(s) at which the outcome is measured: TEXT (max 255 chars)
      }
    ],
    "secondaryOutcomes": [ \\ Conditional. List of secondary outcome measures for additional study assessments: ARRAY of OBJECT
      {
        "measure": \\ Name of the secondary outcome measure: TEXT
        "description": \\ Description of the metric used to characterize the secondary outcome measure: TEXT
        "timeFrame": \\ Time point(s) at which the outcome is measured: TEXT
      }
    ],
    "otherOutcomes": [ \\ Optional. List of other pre-specified outcome measures (excluding post-hoc measures): ARRAY of OBJECT
      {
        "measure": \\ Name of the pre-specified outcome measure: TEXT
        "description": \\ Description of the metric used to characterize the outcome measure: TEXT
        "timeFrame": \\ Time point(s) at which the outcome is measured: TEXT
      }
    ]
  }
},
```

```

Figure 19: Protocol Section - Outcomes Module

```

# CONTEXT #
You are tasked with analyzing clinical trial study reports or papers to extract specific information as structured data. Omit any fields if the corresponding data does not exist in the target articles. Do not invent or assume information that is not present in the given content.
# PAPER CONTENT #
{{pmc_text}}


# TARGET #
Extract the eligibilityModule of the study. It specifies the eligibility criteria for participating in this clinical trial.
# RESPONSE #
The JSON response must adhere to the following structure, where each field is annotated with its expected type.
Ensure that the output is a valid and correctly formatted JSON string.
Format:
```json
{
  "eligibilityModule": {
    "eligibilityCriteria": \\ Inclusion and exclusion criteria for participant selection, formatted as a bulleted list under respective headers: TEXT (max 2000 chars)
    "healthyVolunteers": \\ Indicates if healthy volunteers without the studied condition can participate: BOOLEAN
    "sex": \\ Eligible participant sex: ENUM (FEMALE, MALE, ALL)
    "minimumAge": \\ Minimum age required for participation, with unit of time: TEXT (Years, Months, Weeks, Days, Hours, Minutes, N/A)
    "maximumAge": \\ Maximum age allowed for participation, with unit of time: TEXT (Years, Months, Weeks, Days, Hours, Minutes, N/A)
    "stdAges": \\ Standardized age categories: ARRAY of ENUM (CHILD, ADULT, OLDER_ADULT)
    "studyPopulation": \\ (Observational studies only) Description of the population source for cohorts or groups: TEXT (max 1000 chars)
    "samplingMethod": \\ (Observational studies only) Method used for sampling: ENUM (PROBABILITY_SAMPLE, NON_PROBABILITY_SAMPLE)
  }
}
```

```

Figure 20: Protocol Section - Eligibility Module

```

# CONTEXT #
You are tasked with analyzing clinical trial study reports or papers to extract specific information as structured data. Omit any fields if the corresponding data does not exist in the target articles. Do not invent or assume information that is not present in the given content.
# PAPER CONTENT #
{{pmc_text}}


# TARGET #
Extract the participantFlowModule of the study. It describes the flow of participants through each stage of the study, including enrollment, allocation, follow-up, and analysis.
# RESPONSE #
The JSON response must adhere to the following structure, where each field is annotated with its expected type.
Ensure that the output is a valid and correctly formatted JSON string.
Format:
```json
{
  "participantFlowModule": {
    "preAssignmentDetails": // Description of significant events after participant enrollment but before group assignment: TEXT (max 500 chars)
    "recruitmentDetails": // Key information about recruitment process: TEXT (max 500 chars)
    "typeUnitsAnalyzed": // Unit of analysis (e.g., "Participants", "Eyes", "Lesions"): TEXT
    "groups": [ // ARRAY of OBJECT - Arms/groups in the study flow
      {
        "id": // Unique group identifier. FG000 is the first group, FG001 is the second, and so on: TEXT
        "title": // Short name of the arm/group: TEXT (max 40 chars)
        "description": // Brief description of the arm/group: TEXT (max 1500 chars)
      }
    ],
    "periods": [ // ARRAY of OBJECT - Time periods in the study
      {
        "title": // Period name (e.g., "Overall Study", "Treatment Phase"): TEXT (max 40 chars)
        "milestones": [ // ARRAY of OBJECT - Key milestones
          {
            "type": // Milestone name (e.g., "STARTED", "COMPLETED"): TEXT (max 100 chars)
            "comment": // Additional information about the milestone: TEXT (max 500 chars)
            "achievements": [ // ARRAY of OBJECT - Numbers for each group
              {
                "groupId": // References a group ID: TEXT (max 500 chars)
                "comment": // Explanation if number differs from expected: TEXT (max 500 chars)
                "numSubjects": // Number of participants: TEXT
                "numUnits": // Number of units if different from participants: TEXT
              }
            ]
          }
        ],
        "dropWithdraws": [ // ARRAY of OBJECT - Reasons for not completing
          {
            "type": // Reason category (e.g., "Adverse Event", "Lost to Follow-up"): TEXT (max 100 chars)
            "comment": // Additional details about the reason: TEXT
            "reasons": [ // ARRAY of OBJECT - Numbers for each group
              {
                "groupId": // References a group ID: TEXT
                "comment": // Additional explanation if needed: TEXT
                "numSubjects": // Number of participants: TEXT
              }
            ]
          }
        ]
      }
    ]
  }
}
```

```

Figure 21: Results Section - Participant Flow Module

```

# CONTEXT #
You are tasked with analyzing clinical trial study reports or papers to extract specific information as structured data. Omit any fields if the corresponding data does not exist in the target articles. Do not invent or assume information that is not present in the given content.

# PAPER CONTENT #
{{pmc_text}}


# TARGET #
Extract the baselineCharacteristicsModule of the study.

# RESPONSE #
The JSON response must adhere to the following structure, where each field is annotated with its expected type.
Ensure that the output is a valid and correctly formatted JSON string.
Format:
```json
{
  "baselineCharacteristicsModule": {
    // Baseline demographic and other initial measures, by arm/group.
    "populationDescription": // Brief reason or explanation if baseline participants differ from the assigned groups.
    "typeUnitsAnalyzed": // (Optional) If units are not participants (e.g., eyes, lesions).
    "groups": [ // ARRAY of OBJECT. Arms/groups for baseline assessment. Must include a "Total" group as the last entry.
      {
        "id": // BG000 is the first group, BG001 is the second, and so on.
        "title": // Short label that identifies the group (e.g., "Placebo," "Treatment A", "Total").
        "description": // Brief explanation of the group's characteristics or interventions.
      }
    ],
    "denoms": [ // ARRAY of OBJECT. Structure for Overall Baseline Measure Data (Row).
      {
        "units": // Unit of measure for the data in this row. Default is "Participants".
        "counts": [ // ARRAY of OBJECT. Each object represents a group and its corresponding count in the same order as the "groups" array.
          {
            "groupId": // References an ID from the "groups" array. (e.g., "BG000," "BG001," "BG002").
            "value": // Number of participants in this group.
          }
        ]
      }
    ],
    "measures": [ // ARRAY of OBJECT. Each baseline or demographic characteristic. Required baseline measures include Age, Sex/Gender, Race, Ethnicity (if applicable), and any other measures.
      {
        "title": // ENUM - See required/optional values below
        // Required1: Age => ENUM("Age, Continuous", "Age, Categorical", "Age, Customized")
        // Required2: Sex/Gender => ENUM("Sex: Female, Male", "Sex/Gender, Customized")
        // Required3 (if possible): Race and Ethnicity => ENUM("Race (NIH/OMB)", "Ethnicity (NIH/OMB)", "Race/Ethnicity, Customized", "Race and Ethnicity Not Collected")
        // Required4 (if possible): Region of Enrollment => ENUM("Region of Enrollment")
        // (Optional): Any other measures
        "description": // Additional descriptive information about the baseline measure
        "populationDescription": // (Optional) If the analyzed population differs from the overall baseline population.
        "paramType": // The type of data for the baseline measure. ENUM("COUNT_OF_PARTICIPANTS", "MEAN", "NUMBER", "MEDIAN", "COUNT_OF_UNITS", "GEOMETRIC_MEAN", "LEAST_SQUARES_MEAN", "LOG_MEAN", "GEOMETRIC_LEAST_SQUARES_MEAN")
        "dispersionType": // Baseline Measure Dispersion/Precision. ENUM("STANDARD_DEVIATION", "FULL_RANGE", "INTER_QUARTILE_RANGE", "NA", "CONFIDENCE_80", "CONFIDENCE_90", "CONFIDENCE_95", "CONFIDENCE_975", "CONFIDENCE_99", "CONFIDENCE_OTHER", "GEOMETRIC_COEFFICIENT", "STANDARD_ERROR")
        "unitOfMeasure": // e.g., "Participants", "years", "kg", etc.
        "denoms": [ // ARRAY of OBJECT. Same structure as "denoms" above, if needed for measure-specific denominators.
          {
            "units": // TEXT
            "counts": [ // ARRAY of OBJECT.
              {
                "groupId": // TEXT
                "value": // TEXT
              }
            ]
          }
        ],
        "classes": [ // ARRAY of OBJECT. Within each measure, define rows or classifications.
          {
            "title": // Baseline RowTitle. (e.g., "Sex: Female, Male")
            "denoms": [ // ARRAY of OBJECT. Same structure as "denoms" above, if needed for class-specific counts.
              {
                "units": // TEXT
                "counts": [ // ARRAY of OBJECT.
                  {
                    "groupId": // TEXT
                    "value": // TEXT
                  }
                ]
              }
            ]
          },
          "categories": [ // ARRAY of OBJECT. Each category is essentially a sub-row under the class.
            {
              "title": // e.g., "Female", "Hispanic or Latino"
              "measurements": [ // ARRAY of OBJECT. Data for each group in this category.
                {
                  "groupId": // e.g., "BG000"
                  "value": // e.g., "53", "64.4"
                  "spread": // e.g., "9.4" (for SD), or omitted if not applicable.
                  "lowerLimit": // e.g., "5.6" Based on Measure Type and Measure of Dispersion (e.g., lower limit of Full Range)
                  "upperLimit": // e.g., "9.9" Based on Measure Type and Measure of Dispersion (e.g., upper limit of Full Range)
                }
              ]
            }
          ]
        }
      }
    ]
  }
}
```

```

Figure 22: Results Section - Baseline Characteristics Module

```

# CONTEXT #
You are tasked with analyzing clinical trial study reports or papers to extract specific information as structured data. Omit any fields if the corresponding data does not exist in the target articles. Do not invent or assume information that is not present in the given content.

# PAPER CONTENT #
{{pmc_text}}


# TARGET #
Extract the outcomeMeasuresModule of the study. It contains the results of primary, secondary, and other outcome measures.

# RESPONSE #
The JSON response must adhere to the following structure, where each field is annotated with its expected type.
Ensure that the output is a valid and correctly formatted JSON string.
Format:
```json
{
  "outcomeMeasuresModule": {
    "outcomeMeasures": [ // ARRAY of OBJECT - All outcome measures with results
      {
        "type": // Outcome type: ENUM ("PRIMARY", "SECONDARY", "OTHER_PRE_SPECIFIED", "POST_HOC")
        "title": // Outcome measure title: TEXT (max 255 chars)
        "description": // Detailed description: TEXT (max 999 chars)
        "populationDescription": // Analysis population if different: TEXT
        "reportingStatus": // Whether data is reported: ENUM ("POSTED", "NOT_POSTED")
        "anticipatedPostingDate": // Expected date if not posted: DATE (format: YYYY or YYYY-MM)
        "paramType": // Type of measure: ENUM ("GEOMETRIC_MEAN", "GEOMETRIC_LEAST_SQUARES_MEAN", "LEAST_SQUARES_MEAN", "LOG_MEAN", "MEAN", "MEDIAN", "NUMBER", "COUNT_OF_PARTICIPANTS", "COUNT_OF_UNITS")
        "dispersionType": // Dispersion/precision type: ENUM ("Not Applicable", "Standard Deviation", "Standard Error", "Inter-Quartile Range", "Full Range", "99% Confidence Interval", "97.5% Confidence Interval", "95% Confidence Interval", "90% Confidence Interval", "80% Confidence Interval", "Other Confidence Interval Level", "Geometric Coefficient of Variation")
        "unitOfMeasure": // Unit of measurement: TEXT
        "calculatePct": // Whether to calculate percentage: BOOLEAN
        "timeFrame": // The description of the time point(s) of assessment must be specific to the outcome measure and is generally the specific duration of time over which each participant is assessed (not the overall duration of the study): TEXT (max 255 chars)
        "typeUnitsAnalyzed": // (Optional) If units are not participants (e.g., eyes, lesions): TEXT (max 40 chars)
        "denomUnitsSelected": // Selected denominator units (e.g., Participants, eyes, lesions): TEXT
        "groups": [ // ARRAY of OBJECT - Study arms/groups
          {
            "id": // Unique group identifier. OG000 is the first group, OG001 is the second, and so on: TEXT
            "title": // Group name: TEXT
            "description": // Group description: TEXT
          }
        ],
        "denoms": [ // ARRAY of OBJECT - Denominators
          {
            "units": // Unit type: TEXT
            "counts": [ // ARRAY of OBJECT
              {
                "groupId": // References a group ID: TEXT
                "value": // Count value: TEXT
              }
            ]
          }
        ],
        "classes": [ // ARRAY of OBJECT - Outcome categories/timepoints
          {
            "title": // Category/timepoint name: TEXT
            "denoms": [ // ARRAY of OBJECT. Similar structure as above.
              {
                "units": // TEXT
                "counts": [
                  {
                    "groupId": // References a group ID: TEXT
                    "value": // TEXT
                  }
                ]
              }
            ],
            "categories": [ // ARRAY of OBJECT - Subcategories
              {
                "title": // Subcategory name: TEXT
                "measurements": [ // ARRAY of OBJECT - Results for each group
                  {
                    "groupId": // References a group ID: TEXT
                    "value": // Result value: TEXT
                    "spread": // Spread (e.g., SD, SE): TEXT
                    "lowerLimit": // CI lower limit: TEXT
                    "upperLimit": // CI upper limit: TEXT
                    "comment": // Explanation for NA values: TEXT
                  }
                ]
              }
            ]
          }
        ]
      }
    ]
  }
}

```

Figure 23: Results Section - Outcomes Measures Module 1

```

    "analyses": [ // ARRAY of OBJECT – Result(s) of scientifically appropriate tests of statistical significance of the primary and secondary outcome measures, if any. Such analyses include: pre-specified in the protocol and/or statistical analysis plan; made public by the sponsor or responsible party; conducted on a primary outcome measure in response to a request made by FDA. If a statistical analysis is reported "Comparison Group Selection" and "Type of Statistical Test" are required. In addition, one of the following data elements are required with the associated information: "P-Value," "Estimation Parameter," or "Other Statistical Analysis."
    {
        "paramType": // Parameter type analyzed: ENUM("Cox Proportional Hazard","Hazard Ratio (HR)","Hazard Ratio, Log","Mean Difference (Final Values)","Mean Difference (Net)","Median Difference (Final Values)","Median Difference (Net)","Odds Ratio (OR)","Odds Ratio, Log","Risk Difference (RD)","Risk Ratio (RR)","Risk Ratio, Log","Slope","Other")
        "paramValue": // The name of the estimation parameter, if "Other" Estimation Parameter is selected.: TEXT
        "dispersionType": // Dispersion type: ENUM ("STANDARD_DEVIATION", "STANDARD_ERROR_OF_MEAN")
        "dispersionValue": // The calculated value for the dispersion of the estimated parameter.: TEXT
        "statisticalMethod": // Method used: TEXT (max 150 chars)
        "statisticalComment": // Additional comments: TEXT
        "pValue": // Calculated p-value given the null-hypothesis: TEXT (max 250 chars)
        "pValueComment": // Additional information about p-value.: TEXT
        "ciNumSides": // CI sides: ENUM ("ONE_SIDED", "TWO_SIDED")
        "ciPctValue": // CI percentage (e.g., "95%"): TEXT
        "ciLowerLimit": // CI lower bound: TEXT
        "ciUpperLimit": // CI upper bound: TEXT
        "ciLowerLimitComment": // Lower limit explanation: TEXT (max 250 chars)
        "ciUpperLimitComment": // Upper limit explanation: TEXT (max 250 chars)
        "estimateComment": // Estimation comment: TEXT
        "testedNonInferiority": // Non-inferiority tested: BOOLEAN
        "nonInferiorityType": // Test type: ENUM ("SUPERIORITY", "NON_INFERIORITY", "EQUIVALENCE", "OTHER", "NON_INFERIORITY_OR_EQUIVALENCE", "SUPERIORITY_OR_OTHER", "NON_INFERIORITY_OR_EQUIVALENCE_LEGACY", "SUPERIORITY_OR_OTHER_LEGACY")
        "nonInferiorityComment": // Test explanation: TEXT (max 999 chars)
        "otherAnalysisDescription": // Other analysis details: TEXT (max 500 chars)
        "groupDescription": // Comparison group details: TEXT
        "groupId": // Groups compared: ARRAY of TEXT
    }
]
}
}
```

```

Figure 24: Results Section - Outcomes Measures Module 2

```

# CONTEXT #
You are tasked with analyzing clinical trial study reports or papers to extract specific information as structured data. Omit any fields if the corresponding data does not exist in the target articles. Do not invent or assume information that is not present in the given content.
# PAPER CONTENT #
{{pmc_text}}


# TARGET #
Extract the adverseEventsModule of the study. It contains information about adverse events including serious adverse events, other adverse events, and mortality data.
# RESPONSE #
The JSON response must adhere to the following structure, where each field is annotated with its expected type.
Ensure that the output is a valid and correctly formatted JSON string.
Format:
```json
{{
  "adverseEventsModule": {{
    "frequencyThreshold": // Threshold for reporting other AEs (e.g., "5%"): TEXT
    "timeFrame": // Time period for AE collection: TEXT (max 500 chars)
    "description": // Additional AE collection details: TEXT
    "allCauseMortalityComment": // Explanation about mortality data: TEXT
    "eventGroups": [ // ARRAY of OBJECT - Study arms/groups for AE reporting
      {{
        "id": // Unique group identifier. EG001 is the first group, EG001 is the second, and so on: TEXT
        "title": // Group name: TEXT (max 1500 chars)
        "description": // Group description: TEXT
        "deathsNumAffected": // Number affected by all-cause mortality: TEXT
        "deathsNumAtRisk": // Number at risk for mortality: TEXT
        "seriousNumAffected": // Number with any serious AE: TEXT
        "seriousNumAtRisk": // Number at risk for serious AEs: TEXT
        "otherNumAffected": // Number with any other AE: TEXT
        "otherNumAtRisk": // Number at risk for other AEs: TEXT
      }}
    ],
    "seriousEvents": [ // ARRAY of OBJECT - Serious adverse events by organ system
      {{
        "term": // AE preferred term: TEXT
        "organSystem": // Organ system category: TEXT
        "sourceVocabulary": // Coding dictionary (e.g., "MedDRA 23.0"): TEXT
        "assessmentType": // Collection method: ENUM (NON_SYSTEMATIC_ASSESSMENT, SYSTEMATIC_ASSESSMENT)
        "notes": // Additional description: TEXT
        "stats": [ // ARRAY of OBJECT - Statistics for each group
          {{
            "groupId": // References an event group ID: TEXT
            "numEvents": // Total number of events: TEXT
            "numAffected": // Number of participants affected: TEXT
            "numAtRisk": // Number at risk: TEXT
          }}
        ]
      }}
    ],
    "otherEvents": [ // ARRAY of OBJECT - Other (non-serious) adverse events
      {{
        "term": // AE preferred term: TEXT
        "organSystem": // Organ system category: TEXT
        "sourceVocabulary": // Coding dictionary: TEXT
        "assessmentType": // Collection method: ENUM (NON_SYSTEMATIC_ASSESSMENT, SYSTEMATIC_ASSESSMENT)
        "notes": // Additional description: TEXT
        "stats": [ // ARRAY of OBJECT - Statistics for each group
          {{
            "groupId": // References an event group ID: TEXT
            "numEvents": // Total number of events: TEXT
            "numAffected": // Number of participants affected: TEXT
            "numAtRisk": // Number at risk: TEXT
          }}
        ]
      }}
    ]
  }}
}},

```

```

Figure 25: Results Section - Adverse Events Module

```

# CONTEXT #
You are tasked with analyzing clinical trial study reports or papers to extract specific information as structured data. Omit any fields if the corresponding data does not exist in the target articles. Do not invent or assume information that is not present in the given content.

# PAPER CONTENT #
{{pmc_text}}


# TARGET #
Extract the moreInfoModule of the study. It contains additional information including limitations and caveats, certain agreements, and point of contact.

# RESPONSE #
The JSON response must adhere to the following structure, where each field is annotated with its expected type.
Ensure that the output is a valid and correctly formatted JSON string.

Format:
```json
{
  "moreInfoModule": {
    "limitationsAndCaveats": {
      "description": // Study limitations and caveats discussed in the paper: TEXT (max 500 chars)
    },
    "certainAgreement": {
      "piSponsorEmployee": // Whether PIs are sponsor employees: BOOLEAN
      "restrictionType": // Type of disclosure restriction: ENUM ("LTE60", "GT60", "OTHER")
      "restrictiveAgreement": // Whether restrictive agreements exist: BOOLEAN
      "otherDetails": // Details if restriction type is OTHER: TEXT
    },
    "pointOfContact": {
      "title": // Contact person's name or title: TEXT (max 255 chars)
      "organization": // Contact organization: TEXT (max 255 chars)
      "email": // Contact email: TEXT (max 255 chars)
      "phone": // Contact phone: TEXT (max 30 chars)
      "phoneExt": // Phone extension: TEXT
    }
  }
}
```

```

Figure 26: Results Section - More Info Module

```

# CONTEXT #
You are tasked with analyzing clinical trial study reports or papers to extract specific information as structured data. Omit any fields if the corresponding data does not exist in the target articles. Do not invent or assume information that is not present in the given content.

# PAPER CONTENT #
{{pmc_text}}


# TARGET #
Extract the following modules from the study:
1. conditionBrowseModule: MeSH condition term mappings
2. interventionBrowseModule: MeSH intervention term mappings

# RESPONSE #
The JSON response must adhere to the following structure, where each field is annotated with its expected type.
Ensure that the output is a valid and correctly formatted JSON string.
Format:
```json
{
  "conditionBrowseModule": {
    "meshes": [ \\ Condition MeSH Terms MeSH terms of Condition/Diseases field
      {
        "id": \\ Condition MeSH ID MeSH ID: TEXT,
        "term": \\ Condition MeSH Term MeSH Heading: TEXT
      }
    ],
    "ancestors": [ \\ Ancestors of Condition MeSH Terms Ancestor (higher level and more broad) terms of Condition MeSH terms in MeSH Tree hierarchy
      {
        "id": \\ Condition Ancestor MeSH ID MeSH ID: TEXT,
        "term": \\ Condition Ancestor MeSH Term MeSH Heading: TEXT
      }
    ],
    "browseLeaves": [ \\ Condition Leaf Topics Leaf browsing topics for Condition field
      {
        "id": \\ Condition Leaf Topic ID: TEXT,
        "name": \\ Condition Leaf Topic Name: TEXT,
        "relevance": \\ Relevance to Condition Leaf Topic: ENUM (LOW, HIGH)
      }
    ],
    "browseBranches": [ \\ Condition Branch Topics Branch browsing topics for Condition field
      {
        "abbrev": \\ Condition Branch Topic Short Name: TEXT,
        "name": \\ Condition Branch Topic Name: TEXT
      }
    ],
    "interventionBrowseModule": {
      "meshes": [ \\ Condition MeSH Terms MeSH terms of Condition/Diseases field
        {
          "id": \\ Condition MeSH ID MeSH ID: TEXT,
          "term": \\ Condition MeSH Term MeSH Heading: TEXT
        }
      ],
      "ancestors": [ \\ Ancestors of Condition MeSH Terms Ancestor (higher level and more broad) terms of Condition MeSH terms in MeSH Tree hierarchy
        {
          "id": \\ Condition Ancestor MeSH ID MeSH ID: TEXT,
          "term": \\ Condition Ancestor MeSH Term MeSH Heading: TEXT
        }
      ],
      "browseLeaves": [ \\ Condition Leaf Topics Leaf browsing topics for Condition field
        {
          "id": \\ Condition Leaf Topic ID: TEXT,
          "name": \\ Condition Leaf Topic Name: TEXT,
          "relevance": \\ Relevance to Condition Leaf Topic: ENUM (LOW, HIGH)
        }
      ],
      "browseBranches": [ \\ Condition Branch Topics Branch browsing topics for Condition field
        {
          "abbrev": \\ Condition Branch Topic Short Name: TEXT,
          "name": \\ Condition Branch Topic Name: TEXT
        }
      ]
    }
  },
  ...
}
```

```

Figure 27: Results Section - Condition Browse Module, Intervention Browse Module