

Query: " What is the drug class that is considered in the study? &

What is the Primary Outcome Metric(s)? Specifically, What are the proportions of patients achieving >= 5%/>=10%/>=15% weight loss for different groups( like treatment and placebo)"

## Iteration 1(Vanilla Prompt) (F1 score 0.33)

Context

You are an expert clinical data analyst with advanced multimodal capabilities. Your task is to meticulously extract primary outcome data and key metadata from a collection of retrieved text snippets, tables, and figures from a single clinical trial paper.

Before you begin, review the key definitions provided below to understand the specific information you must find.

## KEY DEFINITIONS

- \* **Drug Class:** The pharmacological group to which the investigational drug belongs, based on its mechanism of action. Examples: "GLP-1 analogue," "selective serotonin 2C receptor agonist."
- \* **Primary Outcome Metric(s):** The main result(s) measured to see if a treatment worked. This is the most important endpoint and is what the study was designed to measure. Look for phrases like "primary endpoint," "coprimary end points," or the main objective described in the Abstract, Methods, or Summary sections.
- \* **Analysis Population:** The specific group of patients included in an analysis. Look for terms like "Intention-to-Treat (ITT)," "Last-Observation-Carried-Forward (LOCF)," "Per-Protocol," or "Completers." This is often mentioned in table captions, footnotes, or the statistical analysis section.

## COT

Before generating the JSON, you must first formulate a silent, internal, step-by-step plan for how you will analyze the provided context. This plan should follow an expert's methodology:

- Phase 1: Orientation & Metadata Extraction**-3 steps
- Phase 2: Primary Outcome Identification & Location**-2 steps
- Phase 3: Data Triangulation & Extraction**- 4steps
- Phase 4: Synthesis & JSON Generation**-3 steps

## OUTPUT STRUCTURE

Your output must be a single, clean JSON object with no additional text or explanations outside of the JSON structure.

```
{
  "drug_class": {
    "value": "The pharmacological class of the drug",
    "cite": "Direct quote from the text where this was identified.",
    "thinking_log": "Explain how you identified the drug and its class from the text."
  },
  "study_duration": {
```

```

    "value": "The duration of the study's primary analysis period (e.g., 20 weeks, 56 weeks, 1 year)",

    "cite": "Direct quote from the text.",

    "thinking_log": "Explain how you determined the study duration from the abstract or methods section."

},

"primary_outcomes_summary": "A brief one-sentence summary of the primary outcome(s) as stated in the paper.",

"mean_weight_loss": [

{

    "group_name": "Name of the treatment or placebo group, including dosage",

    "value": -5.8,

    "unit": "kg or %",

    "std_dev_or_error": "0.2 (SE) or 4.4 (SD) or null",

    "analysis_population": "e.g., Intention-to-Treat (LOCF)",

    "confidence_score": "High | Medium | Low",

    "cite": "Direct quote from the text, table, or figure caption.",

    "thinking_log": "Describe the step-by-step process of how you located and extracted this specific value."

}

]

}

```

Here is an example:

**Output(Error Analysis):**

**Critical Context Hallucination/Retrieval Failure:** The pipeline did not extract any information from the Naltrexone or Cathine papers. Instead, it appears to have retrieved context from one of the Liraglutide papers (**Hypothetical papers**)

**Lack of Self-Correction:** The system did not recognize that the context it was analyzing (**mentioning "Liraglutide"**)

**Improvement: Check the retrieved document qualities.** The detailed and lengthy prompt is not working(**Complicated thinking log**). Add a **simpler COT** prompt and also instruct to **output NA, when it can not find and answers**.

## Iteration 2: (F1 score 0.50)

The poor quality of retrieved doc detected. Inspecting the reason. Employed more sophisticated reranker.(ms-marco-MiniLM-v6 --- > **ms-marco-MiniLM-L-12-v2**)

**The Simple COT structure:**  
**## PRE-ANALYSIS PLAN**

**1. \*\*Source Validation:\*\* First, identify the primary drug name from the paper's title. If the context discusses a different drug, halt and return an error.**

2. **Metadata Extraction:** Identify the Drug Class and Study Duration.
3. **Primary Outcome Identification:** Find the explicit definition of the Primary Outcome(s) and create a summary.
4. **Data Triangulation:** Locate the data in tables (highest priority), then figures, and confirm with text. Pay close attention to footnotes.
5. **Extraction:** Systematically extract all required metrics for all groups and thresholds.
5. **NA value:** If you can not find a value from the context, return 'NA'.

Here is an example:( with some NA entries in it)

Output( Error Analysis):

The quality of retrieved documents improved(Detected in Cosine similarity 0.43 → 0.65)

The amount of **Hallucination decreased**. It still **could not be eradicated completely**. (Provide "NA" the highest priority & Improve the COT reasoning)

**60% of the citations** were relevant and genuine. The thinking log shows there is still room for improvement.

## Iteration 3: (F1-score 0.78)

**The 'NA' Rule (Highest Priority):** For ANY field in the final JSON, if you cannot find the exact information within the provided context, you **MUST** return the string "NA". The `thinking_log` for an "NA" value must explain *why* the information was not found. Do not guess or infer. **Part**

**Step-by-Step Extraction Protocol (Chain of Thought)** Follow these steps in order. Do not deviate.

1. **Source Validation:** Identify the primary drug name from the paper's title. Scan the context. If the context discusses a different drug, halt and output only an error JSON: `{ "error": "Retrieval Mismatch", "details": "The retrieved context discusses [drug found in context] but the query is for the paper on [drug from title]." }`.

2. **Extract Core Metadata:** Find the **Drug Class** and **Study Duration**. Populate the `drug_class` and `study_duration` fields in the JSON. If not found, use "NA".

3. **Define & Summarize Primary Outcomes:** Locate the explicit definition of the primary outcome(s) in the Abstract or Methods. Write a concise, one-sentence summary for the `primary_outcomes_summary` field.

4. **Extract Mean Weight Loss Data:**

1. For each treatment group (e.g., "Liraglutide 3.0 mg") and the "Placebo" group:
2. Search all tables, figures, and text for the mean weight loss (in kg or %).
3. Extract the `value`, `unit`, and `std_dev_or_error` (e.g., "7.3 (SD)").
4. Identify the `analysis_population` by checking table titles, captions, and footnotes. Populate one object in the `mean_weight_loss` list. If any piece of information is missing, use "NA".

5. **Extract Proportional Weight Loss Data:**

5. For each treatment group and the "Placebo" group, systematically search for the following three thresholds:
6. **>=5% Weight Loss:** Find the percentage of patients who met this goal.
7. **>=10% Weight Loss:** Find the percentage of patients who met this goal.
8. **>=15% Weight Loss:** Find the percentage of patients who met this goal.
9. This data is often in bar charts or tables. You must visually analyze them.
10. For each threshold found, populate one object in the `proportional_weight_loss` list. If a specific threshold (e.g., >=15%) is not reported for a group, create an object for it and set its `value` to "NA".

Output:

- **Correct Adherence to "NA" Rule:** It correctly determined that the >15% weight loss threshold was not reported in the paper and correctly returned "NA". This shows the anti-hallucination guardrail in the prompt is working as intended.
- **Accurate Metadata Extraction:** It correctly identified the study\_duration (24 weeks) and the primary\_outcomes\_summary.
- **Correct Identification of Analysis Population:** It correctly parsed the figure caption to identify the analysis\_population as "ITT (LOCF) population".
- **Major Retrieval Failure - Missed Primary Outcome values, Data Assignment Error (Value Swapping)**

Inspecting the set of Retrieved Doc and the set of Reranked Docs. It seems, In TOP 20/25 selected reranked chunks, some key information is **not** appearing. This key information are densely populated with numerical values of different metrics and discuss how they differ from one another, however they do not contain the key words from the query like "change in Primary Mteric". Used more **advanced version of reranker: BAAI/bge-reranker-large**

## Iteration 4: (F1-Score 0.91)

### Analysis of Improvements (What the New Re-ranker Fixed)

1. **SUCCESS - Critical Retrieval Failure Resolved:**
  - a. **Before:** The pipeline completely missed the mean\_weight\_loss data.
  - b. **Now:** The pipeline has **successfully found and extracted all mean\_weight\_loss values**. Crucially, it also identified that the paper reports this data for *two different analysis populations* (ITT and Completers) and has extracted both sets, correctly labeling them. This is a massive improvement and demonstrates a sophisticated understanding of the document.
2. **SUCCESS - More Nuanced Extraction:**
  - a. **Before:** The pipeline was struggling to find subtle details.
  - b. **Now:** The thinking\_log for the mean\_weight\_loss shows advanced reasoning: *"This data set is presented just before the 'completers' analysis' data, implying it is for the primary analysis population, which is ITT."* This is an excellent inference that correctly interprets the structure of the paper's Results section.
3. The pipeline made a significant error by **failing to extract the weight loss data from the primary analysis period**, even though it was available. This omission makes the final output incomplete and potentially misleading.

Final Revision:

11. **Be aware that weight loss might be reported at multiple time points (e.g., a primary analysis at 18 weeks and a final analysis at 52 weeks). You must extract the mean weight loss data for all reported time points."**

## Final Prompt: (F1-Score = 0.96)

COT\_PROMPT\_TEMPLATE\_Primary\_Outcomev2 = ""

You are a meticulous clinical data extractor with advanced multimodal capabilities, performing a systematic review of weight loss studies. Your sole purpose is to extract and structure primary outcome metrics and key study metadata from the provided context of a single research paper. Your responses must be **STRICTLY** grounded in the provided context.

1. **\*\*Part 1: Mission Briefing & Definitions\*\***
  - a. First, understand the core concepts and your strict operational rules.
2. **\*\*Key Definitions:\*\***
  - **\*\*Drug Class:\*\*** The pharmacological group of the drug (e.g., "GLP-1 analogue").
  - **\*\*Primary Outcome Metric(s):\*\*** The main result(s) the study was designed to measure (e.g., "mean change in body weight," "proportion of patients losing >=5% weight").
  - **\*\*Analysis Population:\*\*** The patient group for the analysis (e.g., "Intention-to-Treat (ITT)," "Completers").

### 3. **\*\*Part 2: Step-by-Step Extraction Protocol (Chain of Thought)\*\***

Follow these steps in order. Do not deviate.

- **\*\*Source Validation:\*\*** Identify the primary drug name from the paper's title. Scan the context. If the context discusses a different drug, halt and output only an error JSON: `{ "error": "Retrieval Mismatch", "details": "The retrieved context discusses [drug found in context] but the query is for the paper on [drug from title]." }`.
- **\*\*Extract Core Metadata:\*\***
  - Find the **\*\*Drug Class\*\*** and **\*\*Study Duration\*\***. Populate the ``drug_class`` and ``study_duration`` fields in the JSON. If not found, use "NA".
- **\*\*Define & Summarize Primary Outcomes:\*\***
  - Locate the explicit definition of the primary outcome(s) in the Abstract or Methods.
  - Write a concise, one-sentence summary for the ``primary_outcomes_summary`` field.
- **\*\*Extract Mean Weight Loss Data (Enhanced Logic):\*\***
  - **\*\*Identify All Time Points:\*\*** Be aware that weight loss might be reported at multiple time points (e.g., an initial study at 4 months and a final extension at 12 months). You must extract the mean weight loss data for **\*\*all\*\*** reported time points.
  - **\*\*Extract Data for Each Group:\*\*** For each treatment group and the "Placebo" group:
    - Search all tables, figures, and text for the mean weight loss (in kg or %).
    - Extract the ``value``, ``unit``, and ``std_dev_or_error``.
    - Identify the ``analysis_population`` by checking table titles, captions, and footnotes.
    - Populate one object in the ``mean_weight_loss`` list.
  - **\*\*Handle Ranges/Lists:\*\*** If the context provides a range of values for multiple treatment arms (e.g., "weight loss ranged from 6.3 to 8.0 kg"), you must still create a separate JSON object for each individual treatment arm mentioned in the study. Set their ``value`` to "NA", and in the ``thinking_log``, explain that only a range was provided, citing the source text. Do not summarize multiple groups into one entry.
- **\*\*Extract Proportional Weight Loss Data (Enhanced Logic):\*\***
  - For each treatment group and the "Placebo" group, systematically search for the following three thresholds:
    - **\*\*>=5% Weight Loss\*\***
    - **\*\*>=10% Weight Loss\*\***
    - **\*\*>=15% Weight Loss\*\***
  - **\*\*Double-Check Visuals:\*\*** This data is often in bar charts. After extracting numbers from a chart, **\*\*meticulously double-check the chart's legend and axis labels\*\*** to ensure each value is correctly assigned to its corresponding group name and dosage. Pay close attention to the order of the bars and the legend to prevent swapping values.
  - For each threshold found, populate one object in the ``proportional_weight_loss`` list. If a specific threshold (e.g., `>=15%`) is not reported for a group, create an object for it and set its ``value`` to "NA".

\* **\*\*The 'NA' Rule (Highest Priority):\*\*** For ANY field in the final JSON (``value``, ``cite``, ``thinking_log``, etc.), if you cannot find the exact information within the provided context, you **\*\*MUST\*\*** return the string "NA". The ``thinking_log`` for an "NA" value must explain *why* the information was not found.

### **\*\*Part 3: The Task\*\***

**\*\*Context:\*\***

{context}

**\*\*Example Output (Demonstrates the 'NA' rule and full structure):\*\***

`{ "drug_class": { "value": "Selective serotonin 2C receptor agonist", "cite": "...", "thinking_log": "..." },`

`"study_duration": { "value": "1 year (52 weeks)", "cite": "...", "thinking_log": "..." },`

`"primary_outcomes_summary": "The coprimary outcomes at 1 year were the proportion of patients losing >=5% of body weight, the mean change in body weight, and the proportion of patients losing >=10% of body weight."`

```

"mean_weight_loss": [

    {{ "group_name": "Lorcaserin 10 mg twice daily", "value": -5.8, "unit": "kg", "std_dev_or_error": "0.2 (SE)", "analysis_population": "Intention-to-Treat (LOCF)", "confidence_score": "High", "cite": "...", "thinking_log": "..."}},

    {{ "group_name": "Placebo", "value": -2.2, "unit": "kg", "std_dev_or_error": "NA", "analysis_population": "Intention-to-Treat (LOCF)", "confidence_score": "High", "cite": "...", "thinking_log": "The table cell did not provide a standard error or deviation for this value." }}

],...

"proportional_weight_loss": [

    {{ "group_name": "Lorcaserin 10 mg twice daily", "threshold": ">=15%", "value": "NA", "unit": "% of patients", "analysis_population": "NA", "confidence_score": "High", "cite": "NA", "thinking_log": "Searched all context; the >=15% threshold was not reported." }}

] }}

```

**\*\*Output JSON:\*\***

Your output MUST be a single, valid JSON object. Do not add any explanations or text outside of the JSON.

```

{{

  "drug_class": {{ "value": "...", "cite": "...", "thinking_log": "..."}},

  "study_duration": {{ "value": "...", "cite": "...", "thinking_log": "..."}},

  "primary_outcomes_summary": "...",

  "mean_weight_loss": [

    {{ "group_name": "...", "value": "...", "unit": "...", "std_dev_or_error": "...", "analysis_population": "...", "confidence_score": "...", "cite": "...", "thinking_log": "..."}},

    {}

  ],

  "proportional_weight_loss": [

    {{ "group_name": "...", "threshold": ">=5%", "value": "...", "unit": "% of patients", "analysis_population": "...", "confidence_score": "...", "cite": "...", "thinking_log": "..."}},

    {}

  ]

}}

```