Paper ML17

Prompting for Precision: Leveraging Agentic Workflows to Accelerate Clinical Trial Document Understanding

Jaya Simha Inampudi, Saama Technologies, Campbell CA, USA

Abstract

Clinical trial success hinges on the thorough comprehension of critical documents like protocols, standards, and Statistical Analysis Plans (SAPs). For statistical and clinical programming teams, mastering this extensive documentation is time-consuming yet essential for daily execution. This paper explores the potential of Large Language Models (LLMs) to accelerate this process. We demonstrate that the key to unlocking this potential lies in sophisticated prompt engineering. By embedding deep contextual knowledge of clinical trial processes into prompts, programmers can query complex documentation with high precision, receiving instant, accurate, and relevant answers. This methodology enables rapid information retrieval, clarifies complex methodologies, and ensures consistent interpretation of standards, resulting in significant efficiency gains and reduced onboarding time. This presentation provides a practical framework and real-world examples for developing effective, context-aware prompts, empowering programming teams to enhance productivity and quality in clinical trial execution.

Keywords: LLM, prompt engineering, SAP, protocols, clinical programming, traceability, governance, multi-agent extraction, coverage, QA

1. Introduction

Clinical development organizations rely on complex, high-stakes documents to define study intent and analysis execution. Protocols define objectives, design, endpoints, assessments, and operational conduct. SAPs define statistical estimands, analysis populations, endpoint derivations, modeling approaches, multiplicity procedures, interim analysis boundaries, missing data handling, safety conventions, and deliverable outputs (TFL shells and graphics standards). Standards documents (e.g., CDISC/SDTM/ADaM conventions, controlled terminology rules, and internal programming standards) constrain how that intent becomes regulated deliverables.

For programmers, the daily reality is not simply “reading the SAP”, it is repeatedly answering operational questions:

- *What is the precise TEAE definition and window?*
- *Which population is primary for each endpoint and which deviations gate PP?*
- *Where are censoring rules for the primary TTE endpoint?*
- *What is the multiplicity hierarchy and how does alpha recycle?*
- *Which TFLs correspond to which analyses and estimands?*

These queries are frequent, distributed across teams, and often answered inconsistently when driven by memory rather than evidence.

LLMs appear to offer an ideal interface: ask a question and receive an answer instantly. However, clinical work has a different acceptance standard than general knowledge tasks. A response is useful only if it is:

1. faithful to the source document,
2. consistent across runs, and
3. supported by traceable evidence.

This paper argues that achieving these properties requires shifting from “prompting as a query” to “prompting as a governed extraction workflow,” where deterministic and LLM components each perform the tasks they are best suited for.

2. Problem Statement: Why Naïve LLM Document Q&A Fails in Regulated Work

2.1 Failure modes in clinical document interpretation

In real-world SAPs and protocols, the same concept may appear across multiple sections and appendices, and terms may be used loosely (e.g., “primary endpoint” vs “primary estimand,” “ITT” vs “Full Analysis Set,” “censoring” described in endpoint section but revised in methods). Naïve LLM usage fails because it:

- **Hallucinates under uncertainty** : fills missing details with plausible clinical conventions.
- **Conflates object types:** endpoint definitions get mixed with estimands; analysis methods are summarized as if they define estimands.
- **Loses provenance:** the answer cannot be tied to where it came from, making QC difficult.
- **Misses appendices/tables:** often underweights shell appendices or table-only definitions.
- **Is non-deterministic:** answers vary, list ordering changes, and terminology shifts between runs.
- **Over-reads:** when fed large context, the model averages across sections and can miss nuance or conflicts.

2.2 Requirements for “LLM-ready” clinical workflows

To be acceptable for programmers, an LLM-assisted system must provide:

- **Scope discipline:** each agent only extracts what it owns; no leakage across domains.
- **Structured outputs:** stable schema blocks rather than free text paragraphs.
- **Evidence linkage:** every extracted object has a narrow reference to the source.
- **Gap surfacing:** missing information becomes an explicit flagged issue, not an invented value.
- **Coverage transparency:** show what sections were mapped to which schema paths.
- **Referential integrity:** endpoints referenced by analyses and TFLs must exist and be linkable.

3. Prompting for Precision: A Governance-First Framework

The proposed framework combines deterministic preprocessing with constrained, role-based LLM extraction. The core idea: **reduce the LLM’s problem** from “understand the entire SAP” to “extract a specific schema slice from a small, routed set of chunks, under strict rules.”

3.1 High-level pipeline components

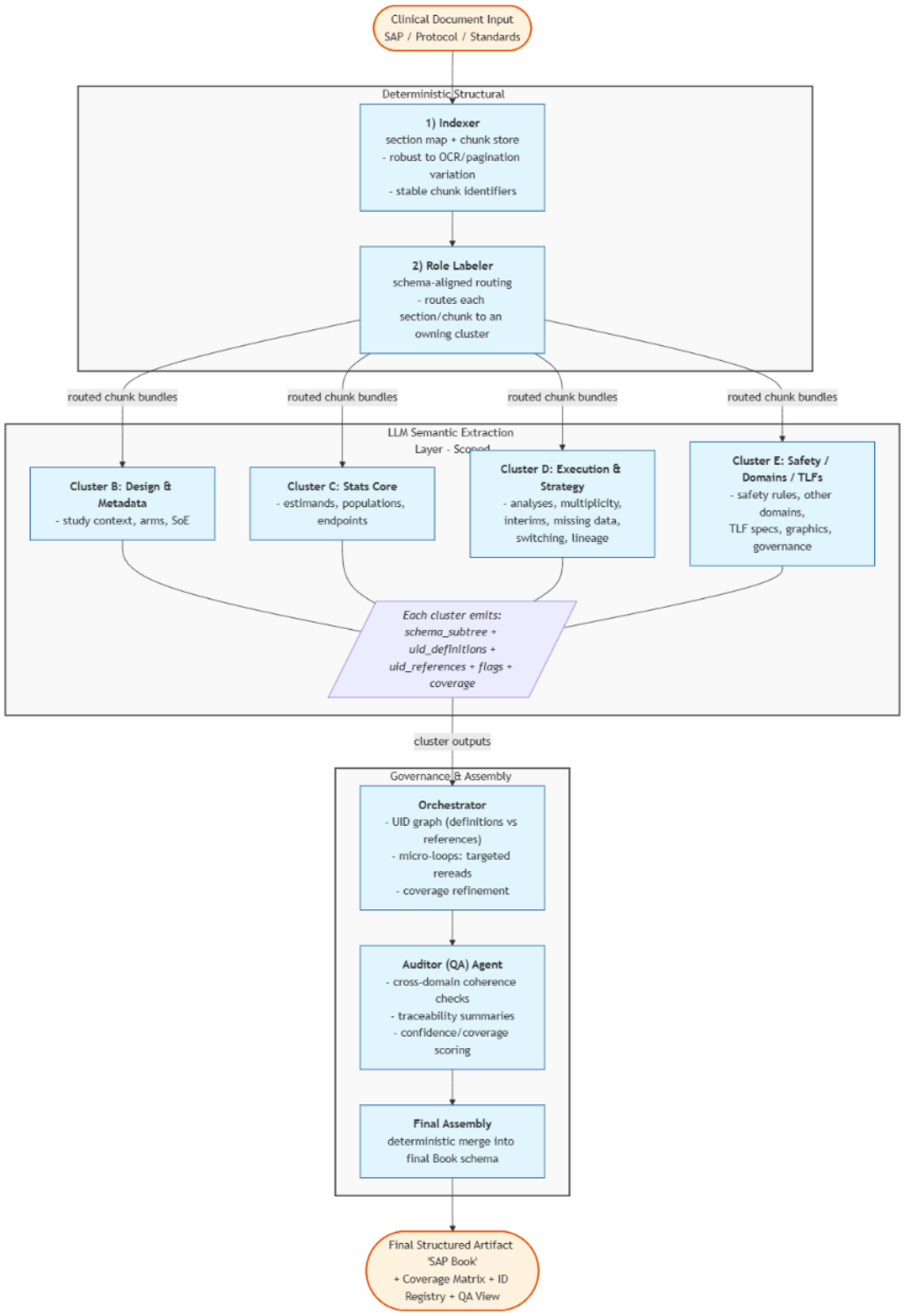
- **Deterministic Indexer:** segments the document into sections and chunks using structure cues; does not rely on complete pagination fidelity.
- **Role Labeler:** assigns chunks/sections to schema-aligned clusters (ownership model).
- **Extraction Clusters (B - E):** LLM agents specialized to emit only their schema subtree, plus UID definitions/references, flags, and coverage updates.
- **Orchestrator:** merges cluster outputs; builds UID graph; runs bounded “micro-loops” for missing definitions; compiles coverage.
- **Auditor (QA) Agent:** cross-checks for coherence, traceability, and coverage, producing QA outputs and an ID registry view.
- **Final Assembly:** deterministic/hybrid merge into the final “Book” artifact (organization-specific schema).

3.2 Why the pipeline is intentionally “not end-to-end LLM”

A fully LLM-driven approach increases unpredictability and token cost. Deterministic indexing and routing keep the LLM focused on semantic interpretation and extraction, where it adds value, while structure and governance remain stable and reproducible.

4. Clean Conceptual Architecture Figure

Below is a conceptual diagram showing the separation of deterministic control and LLM semantic extraction.



5. Deterministic Indexing and Chunking (Why it Matters)

5.1 Structural segmentation as a prerequisite for precision

SAPs are long and unevenly structured: title pages, amendments, SoE tables, appendices of shells, and sometimes multiple TOCs. Indexing serves several purposes:

- **Bounds the LLM context** to relevant material.
- **Enables routing** by section/topic.
- **Provides stable references** for evidence linkage.
- **Supports coverage reporting** (what was read vs ignored).

5.2 OCR-agnostic and pagination-agnostic principles

In practice, some PDFs have inconsistent headers, missing printed page numbers, or OCR artifacts (broken ligatures, hyphenation, character substitutions). An effective indexer must be resilient to:

- missing or inconsistent page numbering,
- slight mismatch between TOC references and body location,
- OCR distortions of section titles and numbering,
- repeated headings (e.g., “Safety” in multiple contexts).

The central design choice is to prioritize **anchor-based segmentation** over page-number segmentation. Anchors are identified using document structure cues (e.g., normalized section IDs + title-like strings) rather than relying on exact page indices.

6. Role Labeling and the “Coverage Map” Ownership Model

6.1 Why ownership matters more than topic labels

Routing by topics such as “Methods” or “Safety” is insufficient because SAP sections often blend content (e.g., endpoint definitions in “Methods,” or population clarifications in appendices). Instead, routing is based on **schema ownership**:

- Each schema block has a primary owner cluster,
- Other clusters may reference it but should not redefine it.

6.2 Practical effect

Ownership prevents “semantic drift.” For instance:

- Cluster D can reference endpoint UIDs but cannot redefine endpoint derivations.
- Cluster E can list which TFLs display an endpoint but does not re-author the endpoint definition.
- The Auditor can highlight conflicts and gaps but does not fabricate missing definitions.

This design reduces inconsistency across teams because extracted objects have one canonical source.

7. Constrained Extraction via Prompt Contracts

7.1 The prompt contract pattern (high level)

Each cluster prompt enforces:

- **Scope:** only allowed schema blocks can be populated.
- **Output schema:** strict YAML/JSON shape with required keys.
- **Null discipline:** absent content must remain null/empty.
- **Evidence requirement:** every object includes a narrow reference (e.g., section ref).
- **Flags:** missing/ambiguous/conflicting information becomes structured flags.

This makes the LLM behave less like a summarizer and more like a schema-constrained extractor.

7.2 Determinism in practice

Determinism is achieved by controlling:

- ordering rules (e.g., sort by UID),
- stable identifiers (UID strategy),
- fixed output schemas and validation outside the LLM.

The system is designed so that when content is ambiguous, it is recorded as ambiguity rather than resolved by assumption.

8. Orchestrator Governance: UID Graph, Micro-loops, and Coverage

8.1 UID graph as a "clinical referential integrity" layer

Clinical SAP interpretation generates entities that must connect cleanly:

- endpoints ↔ estimands ↔ analyses ↔ outputs (TFLs/figures)
- populations used across multiple endpoints and analyses
- interim looks tied to specific analyses
- multiplicity hierarchies referencing endpoints or estimands

The UID graph consolidates:

- which entities are defined,
- where they are referenced,
- what conflicts exist.

This converts a qualitative review problem into a structured, auditable one.

8.2 Micro-loops as bounded, targeted remediation

When an entity is referenced but missing a definition, the orchestrator triggers a targeted reread:

- selects likely owner cluster,
- sends only a small set of candidate chunks,
- requests: define if present; otherwise return gap.

Micro-loops are bounded to avoid infinite loops and to keep costs predictable.

8.3 Coverage refinement

Coverage is tracked at section granularity:

- section → schema paths populated

- section → status: mapped / gap / ignored

This is useful for QC, onboarding, and audit readiness because it is clear what parts of the document were used to populate the structured artifact.

9. Carefully Non-Replicable "Before/After" Examples

The examples below are illustrative.

Example 1 , TEAE Definition Query (Safety)

Before (naïve chat-with-document):

> **User:** What is the TEAE definition in this SAP?
> **Model answer (typical failure pattern):** TEAEs are adverse events occurring from first dose until 30 days after last dose, including events that worsen from baseline.

Issues:

- The model selected "30 days" without evidence.
- No source reference.
- "Worsen from baseline" may be defined differently in the SAP.

After (precision-governed extraction style):

```
safety.teae_definition:
  description: "Treatment emergent adverse events are defined relative to the date of first dose [Week 0 (Visit 3) unless indicated otherwise]."
  lag_time: null
  start_rules: "events with a start date that is equal to or greater than the date of first dose; events that start prior to the date of first dose and worsen after that date; events that start and resolve prior to the date of first dose, but then recur after that the date of first dose."
  stop_rules: null
  worsening_baseline_logic: "events that start prior to the date of first dose and worsen after that date"
  sap_ref: "Sec 11.2"
flags:
  - type: "GAP_SAFETY"
    code: "MISSING_TEAE_LAG_TIME"
    schema_path: "safety.teae_definition.lag_time"
    sap_ref: "Sec X.Y (Safety Definitions)"
    notes: "TEAE lag time after last dose is not explicitly stated in the routed safety text."
```

What changed:

- The system refuses to invent "30 days."
- It returns a structured partial extraction and creates a clear gap for follow-up.

Example 2 , Endpoint ↔ Analysis ↔ TLF Alignment

Before (naïve summarization):

> “Primary endpoint OS is analyzed using Cox model and displayed in Table 14.1.”

Issues:

- “Table 14.1” may not be correct, and OS may be displayed in a figure, not a table.
- Missing population context.
- Missing evidence references.
- No link structure to validate completeness.

After (governed, linked output view):

```
endpoints.items[0]:
  endpoint_uid: "END-EXAMPLE-01"
  name: "Overall Survival"
  class: "Primary"
  type: "TTE"
  definition: "Time from randomization to death from any cause."
  analysis_population_uids: ["POP-EXAMPLE-01"]
  sap_ref: "Sec A.B (Endpoints)"

analyses.items[0]:
  analysis_uid: "ANA-EXAMPLE-01"
  name: "Primary OS model"
  endpoint_uid: "END-EXAMPLE-01"
  population_uid: "POP-EXAMPLE-01"
  model:
    method: "Cox proportional hazards"
    formula: null
  sap_ref: "Sec C.D (Primary Analysis)"

tlf_specifications.items[0]:
  tlf_uid: "OUT-EXAMPLE-01"
  output_id: "Figure F.X"
  type: "Figure"
  title: "Kaplan–Meier Plot of Overall Survival"
  population_uid: "POP-EXAMPLE-01"
  endpoint_uids: ["END-EXAMPLE-01"]
  analysis_uids: ["ANA-EXAMPLE-01"]
  sap_ref: "Annex Z (Shells)"

uid_references:
  - uid: "END-EXAMPLE-01"
    category: "endpoint"
    used_in_path: "analyses.items[0].endpoint_uid"
    sap_ref: "Sec C.D"
  - uid: "ANA-EXAMPLE-01"
    category: "analysis"
    used_in_path: "tlf_specifications.items[0].analysis_uids[0]"
    sap_ref: "Annex Z"
```

What changed:

- The system produces a consistent link graph across endpoint, analysis, and output.
- Each object is anchored to a location (sap_ref).
- QC can now automatically detect if an output references an undefined analysis or endpoint.

10. Implementation Strategy for Clinical Programming Organizations

10.1 Adoption path

A practical rollout typically follows stages:

1. **Read-only Q&A with evidence** (structured answers + references)
2. **Schema extraction for a subset** (e.g., endpoints + populations + TEAE rules)
3. **End-to-end "Book" assembly** with coverage + QA views
4. **Integration** with downstream workflows (spec authoring, QC checklists, TFL libraries)

10.2 Guardrails and governance

To keep the solution safe and credible:

- Enforce schema validation outside the LLM,
- Preserve all flags and gaps as first-class outputs,
- Treat unresolved conflicts as visible issues, not silent overwrites,
- Log coverage and UID consistency to support QC.

10.3 Security and compliance considerations

Organizations should ensure:

- access-controlled document inputs,
- retention policies aligned to SOPs,
- clear separation of "assistive extraction" vs "regulatory decision making,"
- auditable logs of what text was processed.

11. Evaluation Framework (Suggested Metrics and Study Design)

11.1 Efficiency metrics

- Median time-to-answer for top recurring SAP questions
- Time to produce baseline structured extraction for a new study
- Reduction in "document re-read cycles" during QC

11.2 Quality metrics

- Evidence coverage rate: proportion of extracted objects with a reference
- Repeat-run stability: structural equivalence across runs
- Conflict rate: number of UID conflicts or ambiguous definitions surfaced

11.3 Onboarding metrics

- Time for a new programmer to independently answer a standardized set of SAP questions
- Reduction in escalations to senior staff for interpretation questions

12. Limitations and Future Work

12.1 Limitations

- Table-heavy content may require specialized table parsing beyond standard text extraction.
- Some SAPs embed key definitions in images or scanned appendices.
- Conflicts between protocol and SAP require cross-document arbitration logic.

12.2 Future work

- Strengthen table understanding with hybrid table extraction pipelines.
- Expand to multi-document harmonization: Protocol ↔ SAP ↔ Standards.
- Add regression harnesses for deterministic re-runs and drift detection.

13. Conclusion

LLMs can materially accelerate clinical trial document understanding, but only when deployed with governance: deterministic indexing, schema-aligned routing, constrained role-based extraction, explicit gap surfacing, and QA oversight. **Prompting for precision** shifts the system from “summarize the SAP” to “produce a traceable, structured interpretation with coverage transparency,” supporting faster execution, improved consistency, and reduced onboarding time, while maintaining the evidence discipline required for regulated clinical workflows.

Additional Examples:

Endpoints Extraction:


```
"endpoints": {
      "present": true,
      "items": [
        {
          "endpoint_uid": "END-ADASCOG-01",
          "endpoint_id": "Primary Endpoint",
          "name": "Alzheimer's Disease Assessment Scale - Cognitive Subscale, total of 11 items [ADAS-Cog (11)]",
          "class": "Primary",
          "type": "Continuous",
          "definition": "Alzheimer's Disease Assessment Scale - Cognitive Subscale, total of 11 items [ADAS-Cog (11)]
at Week 24",
          "derivation_rules": null,
          "baseline_definition": null,
          "censoring_rules": {
            "description": null,
            "event_vs_censor_hierarchy": null
          },
          "adjudication": null,
          "assessment_source": null,
```

```
        "analysis_population_uids": [
          "POP-ITT-01",
          "POP-EFF-01"
        ],
        "sap_ref": "Sec 3.2.1"
      },
      {
        "endpoint_uid": "END-CIBIC-01",
        "endpoint_id": "Primary Endpoint",
        "name": "Video-referenced Clinician's Interview-based Impression of Change (CIBIC+)",
        "class": "Primary",
        "type": "Ordinal",
        "definition": "Video-referenced Clinician's Interview-based Impression of Change (CIBIC+) at Week 24",
        "derivation_rules": null,
        "baseline_definition": null,
        "censoring_rules": {
          "description": null,
          "event_vs_censor_hierarchy": null
        },
        "adjudication": null,
        "assessment_source": null,
        "analysis_population_uids": [
          "POP-ITT-01",
          "POP-EFF-01"
        ],
        "sap_ref": "Sec 3.2.1"
      },
      {
        "endpoint_uid": "END-NPI-01",
        "endpoint_id": "Secondary Endpoint",
        "name": "Revised Neuropsychiatric Inventory (NPI-X)",
        "class": "Secondary",
        "type": null,
        "definition": "Revised Neuropsychiatric Inventory (NPI-X) to assess the dose-dependent improvement in
behavior",
        "derivation_rules": "Improved scores on the Revised Neuropsychiatric Inventory (NPI-X) will indicate
improvement in these areas.",
        "baseline_definition": null,
        "censoring_rules": {
          "description": null,
          "event_vs_censor_hierarchy": null
        },
        "adjudication": null,
        "assessment_source": null,
        "analysis_population_uids": [],
        "sap_ref": "Sec 3.1.2"
      }
    ],
    "additional_endpoint_properties": [
      {
        "label": "CIBIC+ Categorical Analysis",
        "value": "At the request of the FDA reviewers, treatments will also be compared for overall differences in the
CIBIC+ by using the CMH test, controlling for site.",
        "sap_ref": "Sec 16.1"
```

```
      },
      {
        "label": "Excluded Endpoints",
        "value": "The following efficacy endpoints will not be used: ADAS-Cog (14) and DAD.",
        "sap_ref": "Sec 15.1"
      }
    ]
  },
```


Population Extraction:


```
"populations": {
    "present": true,
    "sets": [
      {
        "population_uid": "POP-ITT-01",
        "name": "Intent-to-Treat Population",
        "short_name": "ITT",
        "definition": "All patients randomized",
        "mapping_to_endpoints": {
          "endpoint_uids": [
            "END-ADASCOG-01",
            "END-CIBIC-01"
          ],
          "endpoint_names": []
        },
        "protocol_deviation_handling": null,
        "sap_ref": "Sec 6"
      },
      {
        "population_uid": "POP-SAF-01",
        "name": "Safety population",
        "short_name": "Safety",
        "definition": "All patients randomized and known to have taken at least one dose of randomized drug",
        "mapping_to_endpoints": {
          "endpoint_uids": [],
          "endpoint_names": []
        },
        "protocol_deviation_handling": null,
        "sap_ref": "Sec 6"
      },
      {
        "population_uid": "POP-EFF-01",
        "name": "Efficacy population",
        "short_name": "Efficacy",
        "definition": "All patients who were randomized and took drug, and have at least 1 post-baseline measure for
both ADAS-Cog and CIBIC+",
        "mapping_to_endpoints": {
          "endpoint_uids": [
            "END-ADASCOG-01",
            "END-CIBIC-01"
          ],
          "endpoint_names": []
        },
```

```
        "protocol_deviation_handling": null,
        "sap_ref": "Sec 6"
      },
      {
        "population_uid": "POP-COMP-01",
        "name": "Completers",
        "short_name": "Completers",
        "definition": "All patients in the efficacy population who completed their Week 24 visit (Visit 12)",
        "mapping_to_endpoints": {
          "endpoint_uids": [],
          "endpoint_names": []
        },
        "protocol_deviation_handling": null,
        "sap_ref": "Sec 6"
      }
    ],
    "additional_population_criteria": [
      {
        "label": "Screen Failures",
        "value": "Patients entered into the study but not assigned to a treatment group are considered to be screen failures. Demographic data for screen failures will be included in the data tabulation datasets, but not in the analysis datasets or in the analyses.",
        "sap_ref": "Sec 6"
      },
      {
        "label": "Randomized",
        "value": "Patients who are enrolled in the study are those who have been assigned to a treatment group. Patients who are entered into the study but fail to meet criteria specified in the protocol for treatment assignment will not be enrolled in the study. Patients are randomly assigned to treatment groups at Week 0 (Visit 3).",
        "sap_ref": "Sec 6"
      }
    ]
  },
```


Missing Data/ Date Imputation Rules:


```
"missing_data": {
    "present": true,
    "rules": [
      {
        "domain": "Adverse Event dates",
        "primary_strategy": "Rule-based imputation using treatment start date relationship",
        "sensitivity_strategies": [],
        "partial_date_imputation_algorithms": "AE start date: Imputed based on relationship to treatment start date (TRTSTD). If AEM missing: no imputation. If AEY<TRTY: TRTSTD+1. If AEY=TRTY and AEM<TRTM: 15MONYYYY. If AEY=TRTY and AEM=TRTM: 01MONYYYY. If AEY=TRTY and AEM>TRTM: 01MONYYYY. If AEY>TRTY: 01JANYYYY or 01MONYYYY depending on month. AE end date: min(date of death if applicable, last day of month) if day missing; min(date of death if applicable, 31DEC) if month and day missing. If imputed start > end, use end as start. Events with start \u2264 cutoff and end missing/after cutoff: impute end as min(cutoff, end of study, death).",
        "notes": "No imputation for missing AE start dates or dates missing year. Imputed dates used for duration calculations but original dates shown in listings. Events continuing beyond cutoff reported as 'continuing'.",
        "sap_ref": "Sec 5.6.1"
      },
      {
```

```
        "domain": "Concomitant medication dates",
        "primary_strategy": "Same as AE start date imputation",
        "sensitivity_strategies": [],
        "partial_date_imputation_algorithms": "Start date follows same conventions as AE date imputation. End dates are not imputed.",
        "notes": "Partial concomitant medication end dates will not be imputed",
        "sap_ref": "Sec 5.6.2"
      },
      {
        "domain": "Prior anti-neoplastic therapies",
        "primary_strategy": "Rule-based imputation relative to study treatment start",
        "sensitivity_strategies": [],
        "partial_date_imputation_algorithms": "Start date: Same as AE imputation except scenario (B) replaced with 'start date of study treatment -1'. End date: min(start date of study treatment, last day of month) if day missing; min(start date of study treatment, 31DEC) if month and day missing. If imputed start > end, use end as start. If both imputed and imputed start > imputed end, use imputed end as start.",
        "notes": null,
        "sap_ref": "Sec 5.6.3"
      },
      {
        "domain": "Post anti-neoplastic therapies",
        "primary_strategy": "Rule-based imputation relative to study treatment end",
        "sensitivity_strategies": [],
        "partial_date_imputation_algorithms": "Start date: max(last date of study treatment + 1, first day of month) if day missing; max(last date of study treatment + 1, 01JAN) if day and month missing. End date: No imputation.",
        "notes": null,
        "sap_ref": "Sec 5.6.3"
      },
      {
        "domain": "Tumor assessment dates",
        "primary_strategy": "Use complete dates only; calculate from available complete dates",
        "sensitivity_strategies": [],
        "partial_date_imputation_algorithms": "All investigation dates must be complete (day, month, year). If incomplete but other dates available: incomplete dates not considered. Assessment date = latest of all investigation dates if CR/CRi/UNK; earliest date if relapsed/no response. If all dates missing day: use first of month. If month missing: use date exactly between previous and following assessment. If no previous/following assessment: not used for calculations.",
        "notes": "Incomplete dates excluded from assessment date calculation when other complete dates available",
        "sap_ref": "Sec 5.6.4"
      },
      {
        "domain": "Relapse or last known date in remission",
        "primary_strategy": "Impute to minimum of assessment date and rule-based date",
        "sensitivity_strategies": [],
        "partial_date_imputation_algorithms": "Missing day: min(date of assessment, 15th day of month and year). Missing day and month: min(date of assessment, 01-Jul of year).",
        "notes": "Used for patients entering secondary follow-up phase while in remission",
        "sap_ref": "Sec 5.6.5"
      },
      {
        "domain": "Death or last known date alive",
        "primary_strategy": "Impute relative to last contact date or assessment date",
        "sensitivity_strategies": [],
```

```
          "partial_date_imputation_algorithms": "Death date: max(full non-imputed last contact date, imputed date). Missing day: 15th day of month and year. Missing day and month: 01-Jul of year. Last known alive: first impute then use for last contact calculation. Missing day: min(date of assessment, 15th day of month and year). Missing day and month: min(date of assessment, 01-Jul of year).",
          "notes": null,
          "sap_ref": "Sec 5.6.6"
        },
        {
          "domain": "Initial diagnosis, first relapse, most recent relapse",
          "primary_strategy": "Impute to minimum of informed consent date -1 and rule-based date",
          "sensitivity_strategies": [],
          "partial_date_imputation_algorithms": "Missing day: min(informed consent date -1, 15th day of month and year). Missing day and month: min(informed consent date -1, 01-Jul of year).",
          "notes": null,
          "sap_ref": "Sec 5.6.7"
        },
        {
          "domain": "Hospitalization dates",
          "primary_strategy": "Same as AE end date imputation",
          "sensitivity_strategies": [],
          "partial_date_imputation_algorithms": "Missing end date or end date after cutoff: follow same conventions as AE end date imputation.",
          "notes": null,
          "sap_ref": "Sec 5.6.8"
        }
      ],
      "additional_missing_data_rules": [
        {
          "label": "Baseline bone marrow tumor burden derivation",
          "value": "If blast cells (%) missing but absolute value available at same visit: convert using formula: % = (absolute/cells counted)*100. If baseline dates differ between aspirate and biopsy: select closest to enrollment. Otherwise select maximum value.",
          "sap_ref": "Sec 5.4"
        }
      ]
    }
```


CONTACT INFORMATION

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

Author Name: Jaya Simha Inampudi
Company: Saama Technologies
Email: jaya.inampudi@saama.com