Integration Of Generative Artificial Intelligence Into The Development Of Statistical Analysis Plans For Patient-reported Outcomes

MSR245

Sameer Mansoori[1], Rashi Rani[2], Akanksha Sharma[1], Rajdeep Kaur[1], Barinder Singh[1], Shubhram Pandey[1]
[1]Pharmacoevidence, Mohali, India, [2]Heorlytics, Mohali, India

PharmacoEvidence — Science-led HEOR & Tech-driven AI

TheoRlytics

INTRODUCTION

- Patient-Reported Outcomes (PROs) are increasingly used in clinical trials and health technology assessments to evaluate treatment benefit from the patient perspective[1]
- Statistical Analysis Plans (SAPs) are critical for ensuring transparency, reproducibility, and regulatory acceptability of PRO analyses[2]
- Development of PRO-specific SAP content is complex and resource-intensive because it requires specification of estimands, handling of intercurrent events, approaches to missing data, multiplicity adjustments, and sensitivity analyses[3,4]
- Regulatory and methodological guidance increasingly emphasize the importance of predefined analytical strategies for PRO endpoints to reduce bias and improve interpretability of trial results[5,6]
- Generative Artificial Intelligence (GenAI) and Retrieval-Augmented Generation (RAG) frameworks have shown potential to improve efficiency and consistency in document generation workflows through context-aware content generation[7]
- Multi-agent AI systems may further enhance SAP development by supporting specialized analytical tasks, iterative refinement, and alignment with methodological guidance[8]

OBJECTIVE

- To develop and evaluate a multi-agent GenAI/RAG framework for automated drafting of PRO-related SAP content, assessing (i) methodological alignment and analytical completeness against expert-developed SAPs, and (ii) efficiency gains relative to conventional development approaches, in the context of clinical and HTA submissions

METHODS

- Regulatory guidance (EMA PRO guidance, ICH E9(R1), NICE PRO guidance), historical SAPs, and PRO methodology literature formed the foundational RAG knowledge corpus
- Context-aware retrieval ensured section-relevant regulatory and methodological content was surfaced at generation time

Figure 1: Overall Approach

1 Framework design	2 Content generation	3 Validation
A structured, prompt-driven multi-agent GenAI framework was developed. Predefined prompts were aligned with regulatory and methodological guidance to ensure consistent, compliant SAP content generation	Thirteen key SAP elements were generated from PRO endpoint definitions through sensitivity analyses and graphical outputs. Iterative feedback loops refined outputs across all development stages	A structured, prompt-driven multi-agent GenAI framework was developed. Predefined prompts were aligned with regulatory and methodological guidance to ensure consistent, compliant SAP content generation

Phase 1 - Framework design

- Six specialized agents designed: Regulatory, Analysis, Longitudinal, Estimand, QC, and cross-sectional consistency
- Predefined prompts aligned to CDISC/ADaM standards and HTA submission requirements

Phase 2 - Content generation

- Thirteen PRO SAP sections generated: endpoint/population definitions, estimand specification, descriptive and longitudinal analyses, subgroup and sensitivity analyses, missing data handling, multiplicity adjustment, responder analyses, and graphical output specifications

Phase 3 - Validation

- Subject matter expert (SME) review assessed methodological alignment, analytical completeness, and internal consistency
- Expert annotations fed back into iterative refinement loops until alignment thresholds were met

Evaluation criteria

- Alignment: AI-generated vs conventionally developed SAP reference documents
- Completeness across all 13 PRO SAP sections
- Efficiency: time-to-draft relative to conventional development

Figure 2: End-to-End pipeline

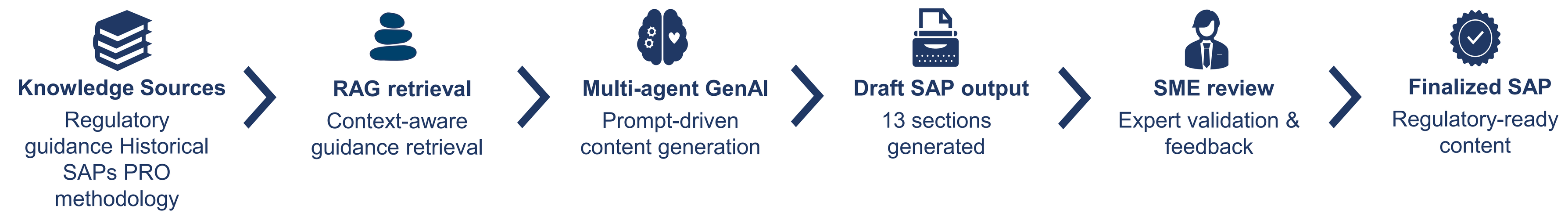


Figure 3: Iterative loop & validation

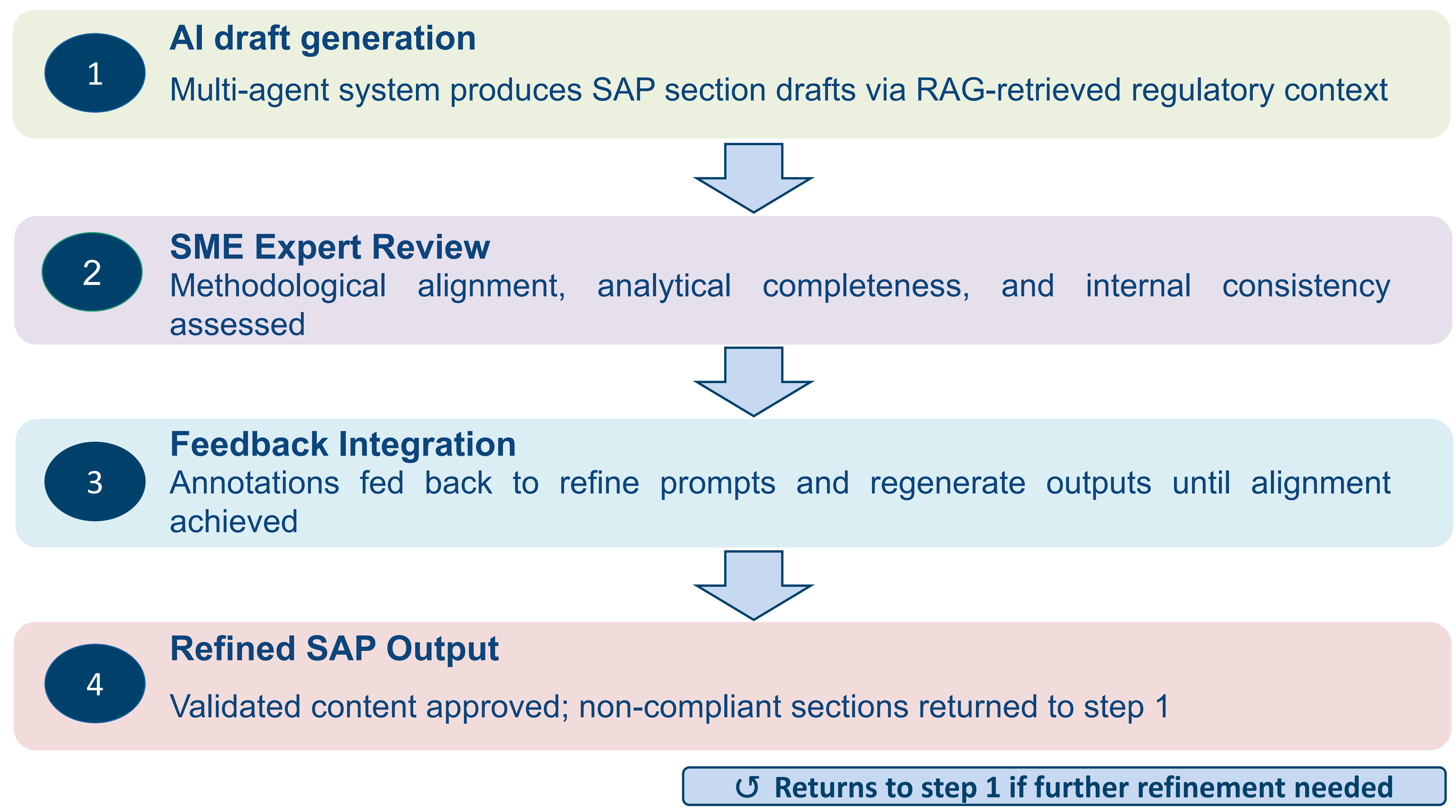


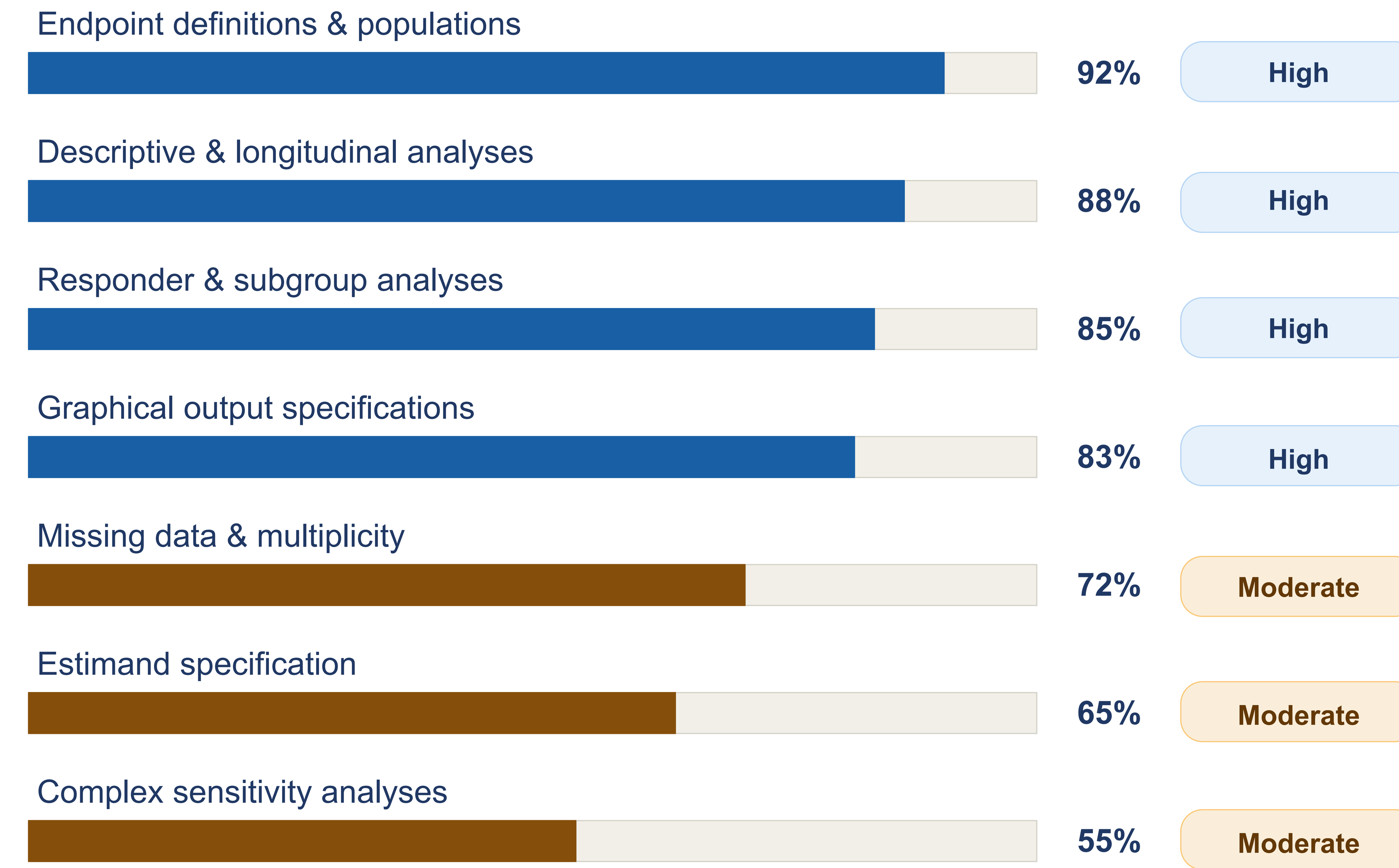
Figure 4: Agent and their roles

- **Regulatory Agent** — Guidance retrieval & regulatory alignment checks
- **Longitudinal Agent** — Repeated measures, MMRM & time-to-event models
- **Analysis Agent** — Statistical methods & primary/secondary analyses
- **Estimand Agent** — Estimand strategy & intercurrent event handling
- **QC Agent** — Cross-section consistency & internal coherence validation

RESULTS

- GenAI-assisted drafting produced SAP content with strong methodological alignment to expert-developed references across standardized sections
- Performance was section-dependent: standardized components aligned well, while estimand specification (65%), sensitivity analyses (55%), and multiplicity handling (72%) required substantial expert input. Alignment results by section are shown in **Figure 5**.

Figure 5: Methodological alignment by sections



9/13	4/13	<50%
sections: high alignment (≥80%)	Sections: expert input needed	time-to-draft vs. conventional development

High alignment achieved

- Endpoint definitions, descriptive/longitudinal, and responder analyses all achieved ≥83% alignment against SME-developed reference SAPs

Reduced development time

- GenAI drafting cut time-to-draft by >50% for well-defined analytical sections relative to conventional approaches

Expert input required

- Estimand specification, sensitivity analyses, and multiplicity sections required substantial SME revision; AI drafts served as starting points rather than final outputs

Iterative refinement necessary

- Mean feedback cycles per section were highest for estimand and sensitivity content, reflecting the nuanced clinical judgement these sections demand

CONCLUSION

- A multi-agent GenAI/RAG framework can produce methodologically aligned PRO SAP content for standardized sections, with >50% reduction in time-to-draft relative to conventional approaches
- Alignment was section-dependent: well-defined analytical sections (endpoint definitions, descriptive and responder analyses) achieved high alignment (≥83%), while complex sections (estimand specification, sensitivity analyses) consistently required expert intervention - establishing a clear division of responsibilities between AI drafting and SME oversight
- These findings support deployment of GenAI-assisted workflows as a drafting aid in clinical and HTA SAP development, with human-in-the-loop validation as a non-negotiable requirement for regulatory submissions
- Evolving regulatory landscape: as PRO guidance continues to develop (particularly post-ICH E9(R1) implementation), RAG corpus maintenance will be critical - static knowledge bases risk generating outdated methodological recommendations

References :
1. FDA. Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. 2009.
2. ICH. ICH E9 Statistical Principles for Clinical Trials. 1998.
3. ICH. ICH E9(R1): Estimands and Sensitivity Analysis in Clinical Trials. 2019.
4. . ISOQOL. Recommendations for analysis of PROs in clinical trials. Quality of Life Research. 2013.
5. EMA. Reflection paper on the use of health-related quality of life measures in oncology studies. 2016.
6. SISAQOL Consortium. Statistical analysis of patient-reported outcomes in cancer randomized trials. Lancet Oncology. 2020.Baio
7. Lewis P, et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. NeurIPS. 2020.
8. Wang L, et al. A Survey of Large Language Model based Multi-Agents. 2024.

Correspondence:
Shubhram Pandey (Shubhram.Pandey@pharmacoevidence.com)

Disclosure:
SM, RR, AS, RK, BS and SP, the authors declare that they have no conflict of interest